Hum Genet (2002) 110 : 451–461
DOI 10.1007/s00439-002-0675-9

ORIGINAL INVESTIGATION

Hong-Wen Deng · Wei-Min Chen · Robert R. Recker

Transmission disequilibrium test with discordant sib pairs when parents are available

Received: 7 August 2001 / Accepted: 28 December 2001 / Published online: 22 March 2002

Abstract The transmission disequilibrium test (TDT) has been employed to map disease susceptibility loci (DSL), while being immune to the problem of population admixture. The customary TDT test (TDT_D) was developed for affected child(ren) and their parents and was most often applied to case-parent trios. Recently, the TDT has been extended to the situations when (1) parents are not available but affected and nonaffected sibs from each family are available, (2) unrelated control-parent trios are available for combined analyses with case-parent trios (TDT_{DC}), and (3) large pedigrees. For many diseases, affected children in the case-parent trios enlisted into the TDT_D have unaffected sibs who can be recruited. We present an extension of the TDT by effectively incorporating one unaffected sib of each of the affected children in the case-parent trios into a single analysis (TDT_{DS}, where DS denotes discordant sib pairs). We have developed a general analytical method for computing the statistical power of the TDT_{DS} under any genetic model, the accuracy of which is validated by computer simulations. We compare the power of the TDT_D, TDT_{DC}, and TDT_{DS} under a range of parameter space and genetic models. We find that the TDT_{DS} is generally more powerful than the TDT_{DC} and TDT_D, particularly when the disease is prevalent (>30%) in the population. The relative power of the TDT_D and the TDT_{DS} largely depends upon the allele frequencies and genetic effects at the DSL, whereas the recombination rate, the degree of linkage disequilibrium, and the marker allele frequencies have little effect. Importantly, the TDT_{DS} not only may be more powerful, it also has the advantage of being able to test for segregation distortion that may yield false linkage/association in the TDT_D.

H.-W. Deng
Laboratory of Molecular and Statistical Genetics,
College of Life Sciences, Hunan Normal University,
ChangSha, Hunan 410081, P. R. China

H.-W. Deng (✉) · W.-M. Chen · R.R. Recker
Osteoporosis Research Center, Creighton University,
601 N. 30th St., Suite 6787, Omaha, NE 68131, USA
e-mail: deng@creighton.edu,
Tel.: +1-402-2805911, Fax: +1-402-2805034601

H.-W. Deng · W.-M. Chen
Department of Biomedical Sciences, Creighton University,
601 N. 30th St., Suite 6787, Omaha, NE 68131, USA

W.-M. Chen
Department of Biostatistics, School of Public Health,
Johns Hopkins University, Baltimore, Md., USA

Introduction

Complex diseases refer to diseases determined by multiple genetic and environmental factors (and potentially their interactions). Linkage disequilibrium (LD) is an important mechanism to identify genes underlying diseases (e.g., Hastbacka et al. 1992, 1994; Xiong and Guo 1997; Deng et al. 2000). Association studies that depend on LD between markers and disease genes have helped to decipher some genetic basis of differential susceptibility to complex diseases (e.g., Feder et al. 1996). Regular association studies, usually, case-control analyses in unrelated cases and controls may suffer inflated type I errors (Chakraborty and Smouse 1988; Lander and Schork 1994; Weir 1996; Spielman and Ewens 1996) that have not been quantified until recently (Deng and Chen 1999; Deng et al. 2001a). In addition, population admixture/stratification may mask or reverse true genetic effects in classical association studies (Deng 2001). Approaches employing nuclear families such as the transmission disequilibrium test (TDT; Spielman et al. 1993) were explicitly proposed to map disease susceptibility loci (DSL). The TDT (Spielman et al. 1993) was developed to control for population admixture/stratification in testing for linkage and/or association between marker loci and DSL.

When parental genotypes are available, the customary TDT analyses only employ affected children and their parents (with at least one parent being heterozygous at the marker locus). Samples commonly used are case-parent trios, i.e., nuclear families with one affected child and both parents. We will denote the customary TDT applied

to case-parent trios as TDT_D, where the subscript D denotes affected children with the disease under study in the trios. Recently, the TDT has been extended to: (1) nuclear families with multiple children (with at least one affected and one unaffected child) where parents are not available (Horvath and Laird 1998; Spielman and Ewens 1998; Boehnke and Langefeld 1998; Knapp 1999); (2) combined samples (TDT_{DC}) of case-parent trios and unrelated control-parent trios (Deng and Chen 2001); (3) general pedigrees (Martin et al. 2000, 2001). It is generally necessary to recruit unrelated control-parent trios in order to rule out segregation distortion and validate significant results in the TDT_D (Spielman et al. 1993). We (Deng and Chen 2001) have previously demonstrated that combined analyses (TDT_{DC}) of case-parent trios and unrelated control-parent trios is not only useful for detecting segregation distortion (which is necessary), but also can increase the mapping power of the TDT (sometimes, dramatically so).

For many diseases, it is generally easy to obtain unaffected sibs for the affected children in the case-parent trios. This is because, for most genetic diseases, the sib recurrence risk is less than 0.5 and, for complex diseases, is usually much lower (Boehnke and Langefeld 1998). If we recruit one unaffected sib for each of the affected children in the case-parent trios, we can form discordant sib-pair-parent tetrads. It is intuitive that such discordant sib-pair-parent tetrads are amenable for TDT type analyses to test for linkage and/or association between markers and DSL, just like the case-parent trios. This is because, if the marker locus under test is a DSL or is linked to and is in LD with a DSL, transmission disequilibrium should occur from heterozygous parents not only to affected children, but also to the nonaffected sibs of the affected children. The direction of the transmission disequilibrium should be opposite if a DSL is involved and should be the same if segregation distortion is involved. Therefore, discordant sib-pair-parent tetrads should be able both to test linkage and/or LD and to test segregation distortion (see Discussion). Compared with the TDT_{DC} analyses that employ case-parent trios and unrelated control-parent trios, genotyping in the TDT_{DS} that is based on discordant sib-pair-parent tetrads is much reduced for the same number of affected and nonaffected children in analyses. This is simply because the discordant sib pairs share both parents. In addition to this advantage of reduced genotyping, we will show, in this paper, that the TDT_{DS} is generally more powerful than the TDT_{DC} and can be frequently more powerful than the TDT_D. Similar to the TDT_{DC}, the TDT_{DS} can also test for the segregation distortion that may plague the results of the TDT_D.

In this article, we will first present the TDT_{DS} test that applies to discordant sib-pair-parent tetrads with at least one parent being heterozygous at the marker locus to test for linkage and/or association. Second, we will develop a general analytical method for computing the statistical power of the TDT_{DS} and the TDT_D. We will validate the accuracy of our power computation method by computer simulations. Finally, under a range of parameter space and genetic models, we will compare the relative powers of the TDT_D, TDT_{DC}, and TDT_{DS}. In the following, we will assume that segregation distortion is absent (see Discussion).

Materials and methods

Statistical tests

For simple illustration, we consider a two-allele-per-locus model at the marker locus having alleles M and m and at a DSL with alleles A and a. This model applies to the data from genetic markers such as single nucleotide polymorphisms and restriction fragment length polymorphisms. For a locus with more than two alleles, such as microsatellite markers, multiple alleles can always be classified into two alleles by designating one (or some) as M and the rest, collectively, as m. In practice, collapsing of multiple alleles into two alleles is not always straightforward, since it is an open question as to which alleles are to be grouped as one allele. Inappropriate collapsing may involve some loss of information. The two-allele model can be extended to account for multiple alleles (Sham and Curtis 1995; Schaid 1996; Spielman and Ewens 1996; Kaplan et al. 1997; Lazzeroni and Lange 1998); this will be pursued in our future studies. The extension can be generally accomplished by testing for global allelic transmission disequilibrium for all alleles instead of two alleles at a time. Therefore, our investigation via the simple two-allele model should be of general significance and forms a basis for future extensions to more complex situations.

In the TDT_D (Spielman et al. 1993) applied to case-parent trios (with at least one heterozygous parent), let T and N_T denote, respectively, the number of times that the marker allele M is transmitted or not transmitted from heterozygous parents to affected children. Under the null hypothesis of no linkage or no LD between the marker locus and a DSL, the statistic

$$\chi^2_{TDT_D} = \frac{(T - N_T)^2}{T + N_T},$$

approximately follows a χ^2-distribution with one degree of freedom (d.f.).

For the TDT_{DC} (Deng and Chen 2001) applied to case-parent trios and unrelated control-parent trios (with at least one parent being heterozygous in each family trio), let $n_{1.}$ denote the total number of M and m alleles transmitted from heterozygous parents to affected children; $n_{2.}$ is similarly defined for unaffected children. Let T_1 and N_{T1} denote, respectively, the numbers of times that the marker allele M is transmitted and not transmitted from heterozygous parents to unaffected children in the control-parent trios. Let $n_{.1}$ denote the number of M alleles transmitted from heterozygous parents to affected and nonaffected children; $n_{.2}$ is similarly defined for the m allele. The total number of alleles transmitted to all children is n_0. Table 1 illustrates the representation of these ns. The statistic

$$\chi^2_{TDT_{DC}} = \frac{n_0\,(T * N_{T1} - T_1 * N_T)^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

approximately follows a χ^2-distribution with 1 d.f. under the null hypothesis of no linkage or no LD between the marker locus and a DSL (Deng and Chen 2001). Note that asterisks in the equations indicate multiplication throughout.

For the TDT_{DS} that employs discordant sib-pair-parent tetrads (with at least one parent being heterozygous), we can construct our $\chi^2_{TDT_{DS}}$ statistic similar to $\chi^2_{TDT_{DC}}$. The difference is that unrelated control-parent trios are replaced by unaffected sibs, one for each of the affected children in the case-parent family trios. The discordant sibs have the same parents. Let $n'_{2.}$ denote the total number of alleles (M and m alleles) transmitted from heterozygous parents to unaffected sibs in the tetrads. Let T_1' and N_{T1}' denote, respectively, the numbers of times that the marker allele M is transmitted and

Table 1 Denotations of the number of alleles transmitted to affected and nonaffected children (*unrelated* numbers in TDT_{DC} analyses for the unrelated control-parent trios, *sib* numbers in TDT_{DS} analyses for the unaffected sibs in the discordant sib pair-parent tetrads)

Number of alleles transmitted			
	M	m	Total
Affected	T	*NT*	$n_{1.}$
Unaffected (unrelated/sib)	$T_1/T_1{}'$	$N_{T1}/N_{T1}{}'$	$n_{2.}/n'_{2.}$
Total (unrelated/sib)	$n_{.1}/n'_{.1}$	$n_{.2}/n'_{.2}$	n_0/n'_0

not transmitted from heterozygous parents to unaffected sibs in the discordant sib-pair-parent tetrads. Let $n'_{.1}$ ($n'_{.2}$) denote the total numbers of M (m) alleles transmitted from heterozygous parents to affected AND nonaffected sibs. The total number of alleles transmitted to all the sibs is n_0'. Table 1 illustrates the representation of n's. The statistic

$$\chi^2_{TDT_{DS}} = \frac{n'_0\left(T * N'_{T1} - T'_1 * N_T\right)^2}{n_{1.}n'_{2.}n'_{.1}n'_{.2}}$$

approximately follows a χ^2-distribution with 1 d.f. under the null hypothesis of no linkage or no LD between the marker locus and a DSL.

Under the alternative hypothesis of linkage and LD between the marker locus and a DSL, $\chi^2_{TDT_D}$, $\chi^2_{TDT_{DC}}$, and $\chi^2_{TDT_{DS}}$ each approximately follows a non-central χ^2-distribution with 1 d.f. and their respective noncentrality parameters being λ_{TDT_D}, $\lambda_{TDT_{DC}}$, and $\lambda_{TDT_{DS}}$. These noncentrality parameters are essential for our analytical approach to compute and compare the statistical power of the TDT_D, TDT_{DC}, and TDT_{DS}. λ_{TDT_D} (for the TDT_D applied to randomly ascertained case-parent trios) and $\lambda_{TDT_{DC}}$ have been developed earlier (Deng and Chen 2001). $\lambda_{TDT_{DS}}$ and λ_{TDT_D} (for the TDT_D applied to the case-parent trios that are ascertained from the discordant sib pair-parent tetrads) will be developed in the next sub-section for our analytical power computation.

Power computation

Let p and q (=*1–p*) denote, respectively, the frequencies of alleles A and a at a DSL. Let f and f' (=*1–f*) denote, respectively, the frequencies of alleles M and m at a marker locus. Let δ denote the LD coefficient for the marker locus and a DSL, and θ denote the recombination rate between the marker locus and a DSL. The population haplotype frequencies are, respectively, $P_{AM}=\delta+pf$, $P_{aM}=f-P_{AM}$, $P_{Am}=p-P_{AM}$, and $P_{am}=1-f-p+P_{AM}$. Let ϕ_{AA}, ϕ_{Aa}, and ϕ_{aa} denote the penetrance (the probability of being affected) of the genotypes AA, Aa, and aa, respectively, at the DSL. This is a general model of within-locus genetic effects at a DSL. Corresponding to the GRR (genotypic relative risk) model that is commonly employed (e.g., Risch and Merikangas 1996), we can define $\phi_{AA}=\gamma_1\phi_{aa}$ and $\phi_{Aa}=\gamma\phi_{aa}$ by utilizing the notations of γ_1 and γ in the GRR model. In the GRR model, on some scales, $\phi_{aa}=1$. The GRR model is useful when computing the power of the TDTs that employ affected children only without consideration of nonaffected children, as the parameter ϕ_{aa} will appear as an independent parameter in the analytical power computation (Deng and Chen 2001). For recessive genetic effects at a DSL, $\gamma=1$; for additive genetic effects, $\gamma_1=2\gamma-1$; for dominant effects, $\gamma_1=\gamma$; and for multiplicative effects, $\gamma_1=\gamma^2$.

Statistical powers of different tests can be investigated and compared by one of the two indices. The first is the probability of rejecting a null hypothesis given a certain sample size at a specified significance level α. The second is the sample size needed to achieve a given power η at a given α. We will choose the second index for this investigation, as the sample size can be unbounded for a statistical power of nearly 100% if the sample exceeds a certain size.

Given the parameters p, f, a, η, θ, δ, ϕ_{AA}, ϕ_{Aa}, and ϕ_{aa}, let us assume that we have N informative nuclear families of the discordant sib-pair-parent tetrads and N case-parent trios are ascertained from the N tetrad families (by only choosing the affected child and the parents from each tetrad family). "Informative nuclear families" refer to those with at least one parent being heterozygous at the marker locus. Only families with one affected child and one unaffected child are sampled, together with their parents. Investigation of the TDT applied to other types of nuclear families, including those specially ascertained through affected status of one or several family members, has been pursued elsewhere (Chen and Deng 2001). For the discordant sib pair-parent tetrads, we will derive the computation methods for and $\lambda_{TDT_{DS}}$ and λ_{TDT_D} for the TDT_{DS} and TDT_D (applied to the case-parent trios ascertained from the tetrads). Let $\Pr(M, Mm \mid D_1=D, D_2=C)$ be the probability that, on condition that one child (D_1) is affected (denoted by D) and the other child (D_2) is nonaffected (denoted by C), a parent heterozygous at the marker locus transmits the M allele to the affected child. $\Pr(m, Mm \mid D_1=D, D_2=C)$ is similarly defined for the transmission of the m allele. The order of children or parents in a tetrad family is not important and is only for notational convenience in the derivation. Among the tetrad families sampled, we denote the affected child as the first child when we derive the probability $\Pr(M, Mm \mid D_1=D, D_2=C)$. The expected numbers of M and m alleles transmitted from marker heterozygous parents to affected children in the tetrad families are, respectively,

$$E(T) = 2N\ \Pr(M, Mm|D_1 = D, D_2 = C), \tag{1}$$

$$E(N_T) = 2N\ \Pr(m, Mm|D_1 = D, D_2 = C). \tag{2}$$

Let $G_i^{P_{Mm}}$ denote the DSL genotype of the parent who is heterozygous at the marker locus. For simplicity, let this parent be the first parent. The two-locus genotype (at the DSL and the marker locus) of this parent $G_i^{P_{Mm}}$ may then be (MA, mA), (MA, ma), (Ma, mA), (Ma, ma), respectively, for i=1, 2, 3, 4, where MA, etc. denote the haplotypes at the marker locus and the DSL. G_j^{Pi} denotes the genotype of the ith parent (i=1, 2) in a tetrad family at the DSL. The superscripts P and O (to appear later) denote the parental and child generation, respectively. Let G_j^{Pi} be AA, Aa, aA, aa, respectively, when j=1, 2, 3, 4 for the ith parent. Then, in the discordant sib-pair-parent tetrads,

$$\begin{aligned}
&\Pr(M, Mm|D_1 = D, D_2 = C)\\
&= \Pr(D_1 = D, D_2 = C, M, Mm) / \Pr(D_1 = D, D_2 = C)\\
&= \sum_{i=1}^{4}\sum_{j=1}^{4} \Pr\left(D_1 = D, M, D_2 = C|G_i^{P_{Mm}}, G_j^{P_2}\right)\\
&\quad \Pr\left(G_i^{P_{Mm}}, G_j^{P_2}\right) / \Pr(D_1 = D, D_2 = C)\\
&= \frac{\sum_{i=1}^{4}\sum_{j=1}^{4} \Pr\left(D_1 = D, M|G_i^{P_{Mm}}, G_j^{P_2}\right)}{\sum_{i=1}^{4}\sum_{j=1}^{4} \Pr\left(D_2 = C|G_i^{P_{Mm}}, G_j^{P_2}\right) \Pr\left(G_i^{P_{Mm}}\right) \Pr\left(G_j^{P_2}\right)}\\
&\quad \frac{}{\Pr\left(D_1 = D|G_i^{P_1}, G_j^{P_2}\right)}\\
&\quad \sum_{i=1}^{4}\sum_{j=1}^{4} \Pr\left(D_2 = C|G_i^{P_1}, G_j^{P_2}\right) \Pr\left(G_i^{P_1}\right) \Pr\left(G_j^{P_2}\right)
\end{aligned} \tag{3}$$

Define the matrices

$$\Phi_D = \begin{pmatrix} \phi_{AA} & \phi_{Aa} \\ \phi_{Aa} & \phi_{aa} \end{pmatrix}, \Phi_C = \begin{pmatrix} 1-\phi_{AA} & 1-\phi_{Aa} \\ 1-\phi_{Aa} & 1-\phi_{aa} \end{pmatrix},$$

$$\text{and } P^{PO} = \begin{pmatrix} 1 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

The numbers in the first row of the P^{PO} are the probabilities that the parents of the genotypes AA, Aa, aA, aa transmit the A allele to a child. The numbers for the second row of the P^{PO} are similarly defined for the transmission of the a allele to a child from the parents of the genotypes AA, Aa, aA, aa, respectively. The probability that the second child is unaffected given that the parents are of the ith and jth genotypes is

$$\begin{aligned}
&\Pr\left(D_2 = C|G_i^{P_1}, G_j^{P_2}\right) \\
&= \sum_{i_2=l}^{2}\sum_{j_2=l}^{2} \Pr\left(D_2 = C, G_{i_2}^{O_2} G_{j_2}^{O_2}|G_i^{P_1}, G_j^{P_2}\right) \\
&= \sum_{i_2=l}^{2}\sum_{j_2=l}^{2} \Pr\left(D_2 = C|G_{i_2}^{O_2} G_{j_2}^{O_2}\right) \Pr\left(G_{i_2}^{O_2}|G_i^{P_1}\right) \Pr\left(G_{j_2}^{O_2}|G_j^{P_2}\right) \\
&= \left(P^{PO}\right)_{.i}^{T} \Phi_C \left(P^{PO}\right)_{.j}
\end{aligned} \tag{4a}$$

where (P^{PO}).i and $(P^{PO})_{.j}$ are, respectively, the ith and jth columns of the matrix P^{PO} and T denotes a matrix transposition. $G_{i_2}^{O_2}$ denotes the DSL allele received by the second (unaffected) child from the parent of the genotype $G_i^{P_1}$ at the DSL, and $G_{j_2}^{O_2}$ is similarly defined for the DSL allele received by the second (unaffected) child from the parent of the genotype $G_j^{P_2}$ at the DSL. Similarly, the probability that a child (the first) is affected given the parental genotypes at the DSL being $G_i^{P_1}$ and G_j^{P2} is

$$\Pr\left(D_1 = D|G_i^{P_1}, G_j^{P_2}\right) = \left(P^{PO}\right)_{.i}^{T} \Phi_D \left(P^{PO}\right)_{.j}. \tag{4b}$$

Define a matrix $P^{PO_M} = \begin{pmatrix} 1/2 & (1-\theta)/2 & \theta/2 & 0 \\ 0 & \theta/2 & (1-\theta)/2 & 1/2 \end{pmatrix}$. The first and second rows of the P^{PO_M} are, respectively, the probabilities that the parents of the two-locus genotypes (heterozygous at the marker locus) (MA, mA), (MA, ma), (Ma, mA), (Ma, ma) transmit M and A alleles (first row) and M and a alleles (second row) to his/her child. Then, the probability that the first parent is a marker heterozygote Mm and the affected child receives the M allele from the first parent given that the parents are of the ith and jth genotypes at the DSL is

$$\begin{aligned}
&\Pr\left(D_1 = D, M|G_i^{P_{Mm}}, G_j^{P_2}\right) \\
&= \sum_{i_1=l}^{2}\sum_{j_1=l}^{2} \Pr\left(D_1 = D, G_{Mi_1}^{O_1} G_{j_1}^{O_1}|G_i^{P_{Mm}}, G_j^{P_2}\right) \\
&= \sum_{i_1=l}^{2}\sum_{j_1=l}^{2} \Pr\left(D_1 = D|G_{Mi_1}^{O_1} G_{j_1}^{O_1}\right) \Pr\left(G_{Mi_1}^{O_1}|G_i^{P_{Mm}}\right) \Pr\left(G_{j_1}^{O_1}|G_j^{P_2}\right) \\
&= \left(P^{PO_M}\right)_{.i}^{T} \Phi_D \left(P^{PO}\right)_{.j}
\end{aligned} \tag{5a}$$

where $G_{Mi_1}^{O_1}$ denotes the allele at the DSL received by the first (affected) child from the the first parent that is of genotype $G_i^{P_{Mm}}$ at the DSL. $G_{j_1}^{O_1}$ denotes the allele at the DSL received by the first (affected) child from the second parent that is of genotype $G_j^{P_2}$ at the DSL. Recall that $G_i^{P_{Mm}}$ denotes the DSL genotype of the first parent heterozygous at the marker locus; the probability that a child is unaffected given that the first parent (P_1) is of genotype $G_i^{P_{Mm}}$ at the DSL and the second parent is of the jth genotype at the DSL is

$$\Pr\left(D_2 = C|G_i^{P_{Mm}}, G_j^{P_2}\right) = \Pr\left(D_2 = C|G_i^{P_1}, G_j^{P_2}\right) \tag{5b}$$

Substituting Equations 4–5 into Equation 3, we have

$$\Pr\left(M, Mm|D_1 = D, D_2 = C\right) = \frac{\left(G^{P_{Mm}}\right)^T \Psi_{Dm} G^{P_2}}{\left(G^{P_2}\right)^T \Psi G^{P_2}}, \tag{6a}$$

where, the ijth element of the 4×4 matrices Ψ_{Dm} and Ψ are, respectively,

$$\begin{aligned}
&\Psi_{Dmij} = \left(P^{PO_M}\right)_{.i}^{T} \Phi_D \left(P^{PO}\right)_{.j} \left(P^{PO}\right)_{.l}^{T} \Phi_C \left(P^{PO}\right)_{.j}, \\
&\Psi_{ij} = \prod_{k=l}^{2} \left(P^{PO}\right)_{.l}^{T} \Phi_k \left(P^{PO}\right)_{.j}, \text{ where } \Phi_1 = \Phi_D \text{ and } \Phi_2 = \Phi_C.
\end{aligned} \tag{6b}$$

$\left(G^{P_2}\right)^T$ is a row vector of the frequencies of the genotypes AA, Aa, aA, and aa, respectively. $\left(G^{P_{Mm}}\right)^T$ is a row vector of the frequencies of the two-locus genotypes (MA, mA), (MA, ma), (Ma, mA), (Ma, ma), respectively. Although not necessary for the validity of the TDT tests, if we assume Hardy-Weinberg equilibrium in the study population (for ease of power computation), we have

$$\begin{aligned}
&\left(G^{P_1}\right)^T = \left(G^{P_2}\right)^T = \left(p^2 \;\; pq \;\; pq \;\; q^2\right), \\
&\left(G^{P_{Mm}}\right)^T = \left(2P_{AM}P_{AM} \;\; 2P_{AM}P_{am} \;\; 2P_{aM}P_{Am} \;\; 2P_{aM}P_{am}\right)
\end{aligned}$$

Similarly, it can be shown that the probability that, in the discordant tetrad families, a parent heterozygous at the marker locus transmits the m allele to the affected child (denoted as the first one for notational convenience) is

$$\Pr\left(m, Mm|D_1 = D, D_2 = C\right) = \frac{\left(G^{P_{Mm}}\right)^T \Psi_{Dm} G^{P_2}}{\left(G^{P_2}\right)^T \Psi G^{P_2}} \tag{6c}$$

where, the ijth element of the 4×4 matrix Ψ_{Dm} is:

$$\Psi_{Dmij} = \left(P^{PO_m}\right)_{.i}^{T} \Phi_D \left(P^{PO}\right)_{.j} \left(P^{PO}\right)_{.i}^{T} \Phi_C \left(P^{PO}\right)_{.j}, \tag{6d}$$

and $P^{PO_m} = \begin{pmatrix} 1/2 & \theta/2 & (1-\theta)/2 & 0 \\ 0 & (1-\theta)/2 & \theta/2 & 1/2 \end{pmatrix}$. The numbers in the first and second rows of P^{PO_m} are, respectively, the probabilities that the parents of the two-locus genotypes (heterozygous at the marker locus) (MA, mA), (MA, ma), (Ma, mA), (Ma, ma) transmit m and A (first row) and m and a (second row) alleles to his/her child.

Equations 6a and 6c can be used to calculate *E(T)* and *E(NT)* (Equations 1 and 2). Hence, under a specified significance level α and a specified statistical power η, the sample size for the $\mathrm{TDT_D}$ applied to the case-parent trios formed from the discordant sib-pair-parent tetrads can be computed from the noncentrality parameter

$$\lambda_{TDT_D} = \frac{[E(T) - E(N_T)]^2}{E(T) + E(N_T)}, \tag{7}$$

by using the numerical procedures adopted and detailed in Deng et al. (2001a) and Deng and Chen (2001).

Let Pr(*M, Mm*|D_1=*C*,D_2=*D*) and Pr(*m, Mm*|D_1=*C*,D_2=*D*) be the probabilities that, conditional on that one child (D_1) is unaffected and the other child (D_2) is affected, a parent heterozygous at the marker locus transmits M and m alleles, respectively, to the unaffected child. It can be shown as above that

$$\Pr\left(M, Mm|D_1 = C, D_2 = D\right) = \frac{\left(G^{P_{Mm}}\right)^T \Psi_{CM} G^{P_2}}{\left(G^{P_2}\right)^T \Psi G^{P_2}} \tag{8a}$$

$$\Pr\left(m, Mm|D_1 = C, D_2 = D\right) = \frac{\left(G^{P_{Mm}}\right)^T \Psi_{Cm} G^{P_2}}{\left(G^{P_2}\right)^T \Psi G^{P_2}} \tag{8b}$$

The matrices Ψ_{CM} and Ψ_{Cm} are computed in the same way as that for the Ψ_{DM} and Ψ_{Dm}, except that Φ_C and Φ_D are exchanged in Equations 6b and 6d. Then, in Table 1,

$$E\left(T_1'\right) = 2N \Pr\left(M, Mm|D_1 = C, D_2 = D\right), \tag{9a}$$

$$E\left(N_{T1}'\right) = 2N \Pr\left(m, Mm|D_1 = C, D_2 = D\right). \tag{9b}$$

The sample size needed under a specified significance level α and a specified statistical power η for the $\mathrm{TDT_{DS}}$ can be computed from the noncentrality parameter

$$\lambda_{TDT_{DS}} = \frac{E\left(T + N_T + T_1' + N_{T1}'\right)\left(E(T)E\left(N_{T1}'\right) - E\left(T_1'\right)E(N_T)\right)^2}{E(T + N_T)E\left(T_1' + N_{T1}'\right)E\left(T + T_1'\right)E\left(NT + N_{T1}'\right)}, \tag{10}$$

by using the numerical procedures adopted and detailed in Deng et al. (2001a) and Deng and Chen (2001). The computation can be implemented easily with the aid of some computer programming. A computer program written in C++ is available from the authors

upon request. It can be shown with some additional algebra that $\lambda_{TDT_{DS}}$ has multiplicative factors of (1/2–θ) and δ so that $\lambda_{TDT_{DS}}$ is zero under the null hypothesis of no linkage or no LD between the marker locus and the DSL. Hence, TDT_{DS} should not suffer inflated type I error rates in the presence of population admixture, as confirmed in our computer simulations.

To validate all the above derivations and our analytical power computation for the TDT_D (applied to the case-parent trios ascertained from the discordant sib-pair-parent tetrads) and the TDT_{DS} from the noncentrality parameters (Equations 7 and 10), we perform computer simulations. The validation of our power computation that is based on the complex analytical derivation by computer simulations is necessary, particularly given that an approximation of the test statistics to the χ^2-distribution is used. In the absence of segregation distortion, random-mating populations are simulated, in which p, γ_1,γ, and ϕ_{aa} are specified together with f, δ, and θ (when the marker locus is not a DSL per se). Note, random mating is not necessary for the validity of the TDT analyses that control for population admixture by nuclear families (Ewens and Spielman 1995). For a desired statistical power η and a specified significance level α, we first compute the sample size (N) needed by our analytical power computation method. Then N discordant sib-pair-parent tetrads are simulated. The TDT_{DS} is applied to the N tetrad families. The simulation and analysis program for this study was developed by the authors and is available upon request. The simulation procedures are relatively straightforward and have been detailed elsewhere (e.g., Deng et al. 2001b; Deng and Chen 2001) and thus will not be elaborated here. Briefly, for the tetrad families, parental genotypes are first simulated based upon the population frequencies of genotypes. For those parents with at least one of whom being heterozygous at the DSL under study, genotypes of children (two from each family) are simulated based upon parental genotypes. The phenotypes are then simulated based upon the genotype-specific penetrances. The families with discordant sib pairs are used for subsequent analyses. In simulations, the random number generator that we used is as given by Park and Miller (1988) and the random number seed is the time to seconds of the computer clock at the time that the simulations are started. The statistical power η' obtained in simulations under the significance level α can be compared with the specified level of η in the analytical power computation. The closer that η' is to η, the more accurate is our analytical power computation. The analytical power computation for the TDT_{DC} was derived and validated previously (Deng and Chen 2001). Once our analytical power computation for the TDT_{DS} and TDT_D is validated by computer simulation, the investigation of the relative power of the TDT_{DS}, TDT_D, and TDT_{DC} will be conducted by our analytical method.

Results

Accuracy of our analytical power computation

Table 2 presents some representative data of our extensive simulation studies for a range of parameters. It can be seen that the sample sizes (N) computed from our analytical method under a specified statistical power (η), if employed in computer simulations, can yield a simulated statistical power (η') that is very close to η. Therefore, the accuracy of our analytical derivation and the power computation for the TDT_{DS} and TDT_D (applied to family trios each with an affected child ascertained from the family tetrads with discordant sibs) is validated by our computer simulations. Simulation results not shown here for other genetic models and parameter values (including those employed in Figs. 1, 2, 3, 4) show a similar accuracy of our analytical power computation to that presented in Table 2.

Validity of the TDT_{DS}

Table 2 presents some results of the simulated significance levels α' under the null hypothesis that the marker locus is not linked to and/or is not in LD with a DSL for the TDT_{DS}. It can be seen that, for various genetic effects at the DSL, the simulated significance level is essentially equal to the specified significance level α=0.01, except the minor differences caused by sampling. Therefore, the TDT_{DS} as a test statistic approximated by a χ^2-distribution with 1 d.f. is valid and robust in that the significance level achieved is the about same as that specified in the TDT_{DS} testing. Data not presented here for other parameter val-

Table 2 The number of nuclear families (N) needed to achieve 90% power η with α=0.01 computed by our analytical methods, and the power (η') obtained by simulations with the sample size N. In the studies for this table, dominant effects are assumed, $\gamma_1=\gamma=2$, and the locus under test is a DSL per se and $f=q$. ϕ_{aa} is the genotypic penetrance for the referent genotype aa. The disease allele A frequency p is 0.1. ϕ is the disease population prevalence and can be computed by the p, γ_1,γ, and ϕ_{aa}. To obtain the power η=90% with the significance level α=0.01 for TDT_{DS} and TDT_D (applied to the family trios selected from the family tetrads), sample size N is calculated from the theoretical noncentrality parameters as indicated in the Methods section. The simulated statistical powers (η') with sample size N are obtained by counting the times that the null hypothesis is rejected in 10,000 repeated simulations performed under the alternative hypothesis as specified; α' is the type one error rate obtained in 10,000 simulations under the null hypothesis that the marker is of no linkage or no LD with a DSL when specifying the significance level α=0.01 by using the χ^2-distributions to the test statistics. TDT_D is the classical TDT applied to family trios each with an affected child selected from the family tetrads recruited for the TDT_{DS}. $TDT_{D'}$ is the classical TDT applied to family trios randomly ascertained. The results given for the $TDT_{D'}$ is obtained from our previous work (Chen and Deng 2001). The *numbers within brackets* are for the individuals needed to be genotyped. The data for PDT is the simulated power of the PDT test of Martin et al. (2000) as reflected by the number of nuclear family tetrads with discordant sib pairs in order to achieve 90% power (η) with α=0.01. Since the power computation and the validity of the PDT and the $TDT_{D'}$ has been substantiated by previous work, their simulated type one error rates and the simulated power (for the $TDT_{D'}$) are not presented here

ϕ_{aa}	$\phi_{AA}=\phi_{Aa}$	ϕ	$TDT_{D'}$	TDT_{DS} (η', α')	TDT_D(η', α')	PDT
0.1	0.2	0.119	376 [1128]	546 (0.902, 0.012) [2184]	373 (0.901, 0.010) [1119]	366 (0.904) [1464]
0.2	0.4	0.238	376 [1128]	384 (0.904, 0.009) [1536]	388 (0.895, 0.009) [1164]	325 (0.900) [1300]
0.3	0.6	0.357	376 [1128]	249 (0.906, 0.009) [996]	410 (0.894, 0.010) [1230]	283 (0.896) [1132]
0.4	0.8	0.476	376 [1128]	140 (0.907, 0.011) [560]	450 (0.899, 0.008) [1350]	222 (0.901) [888]

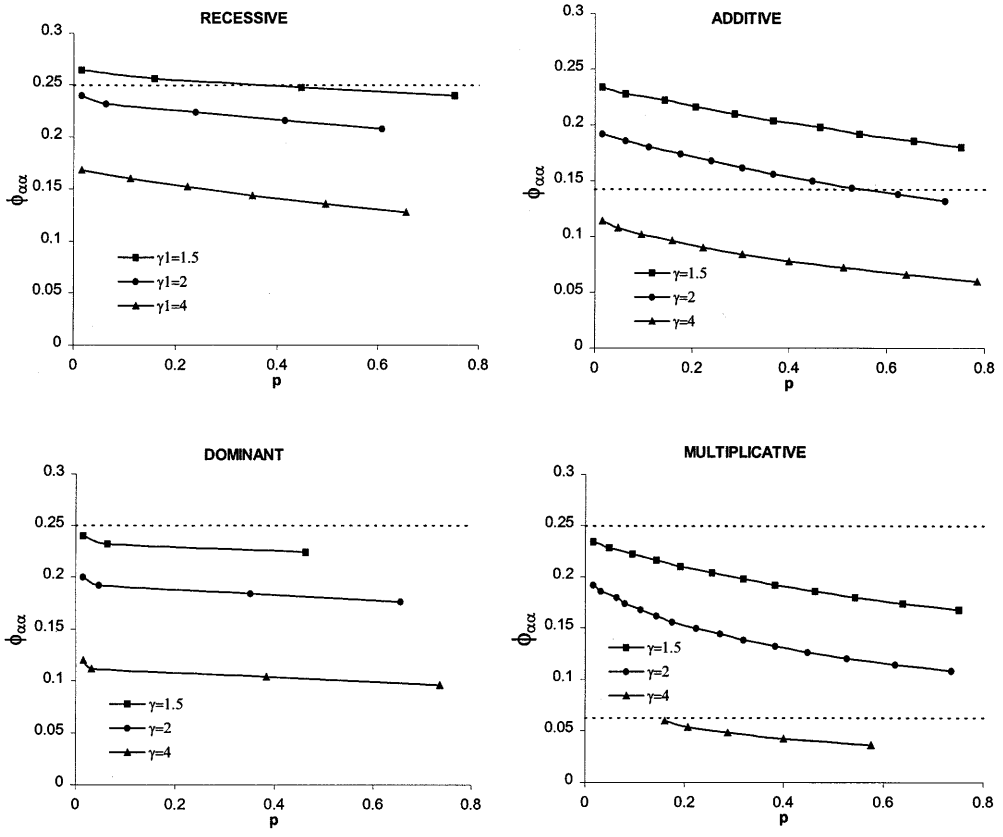

Fig. 1 The regions in the two dimensional parameter (p and ϕ_{aa}) space in which the TDT_{DS} is more powerful than the TDT_D when the marker locus is a DSL per se and $p=f$. In the comparison, η=90% and $\alpha=10^{-7}$, although the choice of η and α is not important for the purpose of comparison of different tests. The *threshold lines* divide the parameter space into two parts. In the parameter space to the *upper right* of the threshold lines, the sample size (N) needed is smaller in the TDT_{DS} than in the TDT_D; thus, the TDT_{DS} is more powerful than the TDT_D in this parameter region. In the parameter space to the *lower left* of the threshold lines, it is the other way around. The *dashed lines* set the upper limit for the maximum values that ϕ_{aa} can take with the constraint that the disease population prevalence ϕ has to be less than 1.0. Under the multiplicative model, there are *two dashed lines* for γ=2 and 4, respectively, and, under the other three models, there is only *one dashed line* for γ=4. In the recessive model, γ_1 is specified, and the penetrance for the genotypes AA, Aa, and aa are, respectively, γ_1 ϕ_{aa}, ϕ_{aa}, and ϕ_{aa}. For the other three models, γ is specified, and γ_1 can be easily inferred from γ and the genetic models under study

ues and genetic models have revealed the same conclusion. In addition, the α' for the TDT_D applied to the family trios (each with one affected child) selected from the family tetrads (each with discordant sib pairs) also confirms to the specified α, so that the classical TDT_D is valid under the ascertainment.

General comparison of the TDT_D and TDT_{DS} tests

It is noted in Table 2 that, for the specified η and a, first, the sample size N required for the TDT_D applied to the case-parent trios ascertained from the discordant sib pair-parent tetrads increases with an increasing ϕ_{aa} and an increasing population disease prevalence ϕ associated with the change of ϕ_{aa}. This is an interesting contrast with the finding (Deng and Chen 2001) that the power of the TDT_D applied to randomly ascertained case-parent trios is not influenced by ϕ_{aa} and the associated change of ϕ. This is largely because the expected frequencies of heterozygous parent does not change with changing ϕ_{aa} in randomly ascertained eligible case-parent trios but decreases in the eligible case-parent trios ascertained from the eligible tetrad families (H.-W. Deng and W.-M. Chen, unpublished). The information for the TDT largely resides in the transmission of alleles from heterozygous parents to children. With an increasing ϕ_{aa}, the TDT_D becomes less powerful with a decreasing number of heterozygous parents in the case-parent trios ascertained from the discordant sib-pair-parent tetrads. However, this is not the case (Deng and Chen 2001; Table 1) when applying the TDT_D to randomly ascertained family trios each with an affected child, a sampling scheme that is probably more often adopted in practice when applying the TDT_D.

Second, the N needed for the TDT_{DS} test decreases with an increasing ϕ_{aa} (and an increasing ϕ that is attributable to the increase of ϕ_{aa}). This may be because, as the disease becomes more prevalent, additional unaffected children become more informative and the contrast between the affected and unaffected children becomes more dramatic (for ϕ<0.5). Third, the power of the TDT_{DS} relative to that of the TDT_D increases with an increasing ϕ_{aa}

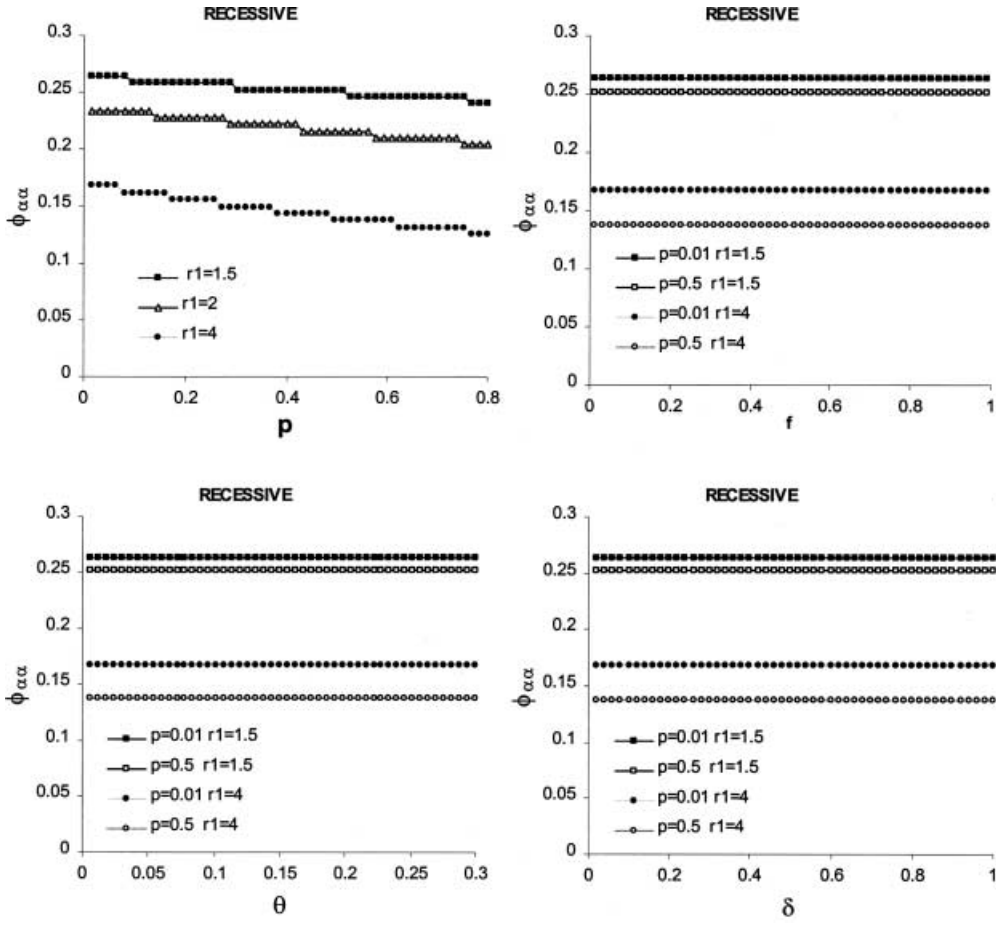

Fig. 2 The parameter space in which the TDT_{DS} is more powerful than the TDT_D when the marker locus is not a DSL. Five parameters p, f, θ, δ, and ϕ_{aa} are investigated. The default parameters are ϕ_{aa}=0.1, f=0.2, and the LD in population δ=0.8 δ_{max}, where δ_{max} is the maximum LD between the marker and the DSL in a population and can be easily shown to be the minimum of pf' and qf. δ varies except in the *lower left* plot where δ is fixed. In the investigation of the allele frequency at the DSL (p), three γ_1 values (1.5, 2, and 4) are investigated, and in other cases, two γ_1 values (1.5 and 4) are investigated in combination with two p values (0.01 and 0.5). Under various models, γ can be inferred easily from the ϕ_{aa} and γ_1 values (see text)

and the associated increasing ϕ. In the investigation for Table 2, the TDT_{DS} is always more powerful than the TDT_D once ϕ_{aa} exceeds 0.2 and ϕ exceeds 0.238. The difference of N for the TDT_D and TDT_{DS} can be so dramatic that the sample size N needed for the TDT_{DS} may be much less than that for the TDT_D, especially when the disease prevalence is large (ϕ>0.36).

Detailed comparison of the TDT_{DS} with the TDT_D and TDT_{DC}, respectively

The relative power and the advantages and disadvantages of the TDT_D (applied to randomly ascertained case-parent trios) and the TDT_{DC} (applied to case-parent trios and unrelated control-parent trios) have been compared in detail previously (Deng and Chen 2001). Therefore, we will focus on the comparison of the relative power of the TDT_{DS} with the TDT_D and TDT_{DC}, respectively. Since the relative magnitudes of the power of the TDT_{DS} and TDT_D have been outlined under some parameters in Table 2, we will focus on this subsection in the parameter regions in which the power of the TDT_{DS} is higher or lower than the TDT_D. Figure 4 gives more data for a comparison of the relative power of the TDT_{DS} and TDT_D.

The comparison is made for the situation when the marker locus is a DSL and p=f (Fig. 1) and for the situation when the marker locus is not a DSL but is linked to and is in LD with a DSL (Fig. 2). When the marker locus is a DSL and p=f (Fig. 1), similar patterns emerge for all the four genetic models investigated. First, there is a continuous threshold line in the parameter space so that, on one side (the upper right side) of the threshold line, the TDT_{DS} is more powerful than the TDT_D in that fewer samples are necessary to reach the same statistical power η for a given significance level α. On the other side (the lower left side) of the threshold line, it is the other way around. Second, ϕ_{aa} is critical in determining the relative power of the TDT_{DS} and the TDT_D. For a p (the frequency of the A allele at the DSL), with an increasing ϕ_{aa}, the power of the TDT_{DS} relative to that of the TDT_D will increase, and after a threshold value ϕ_{aa}, the power of the TDT_{DS} will be larger than that of the TDT_D. The threshold values of ϕ_{aa} decrease with the increasing within-locus relative genetic effects as reflected by γ_1 or γ. Third, p also plays some role in determining the relative power of the TDT_{DS} and TDT_D tests, although the role is not as important as that of ϕ_{aa}, particularly under the dominant genetic model. Similarly, for some ϕ_{aa} values, with an increasing p, the power of the TDT_{DS} relative to that of the TDT_D will increase, and after a threshold value p, the power of the TDT_{DS} will be larger than that of the TDT_D. Again, the threshold p value decreases with increasing within-locus relative genetic effects as reflected by γ_1 or γ. Therefore,

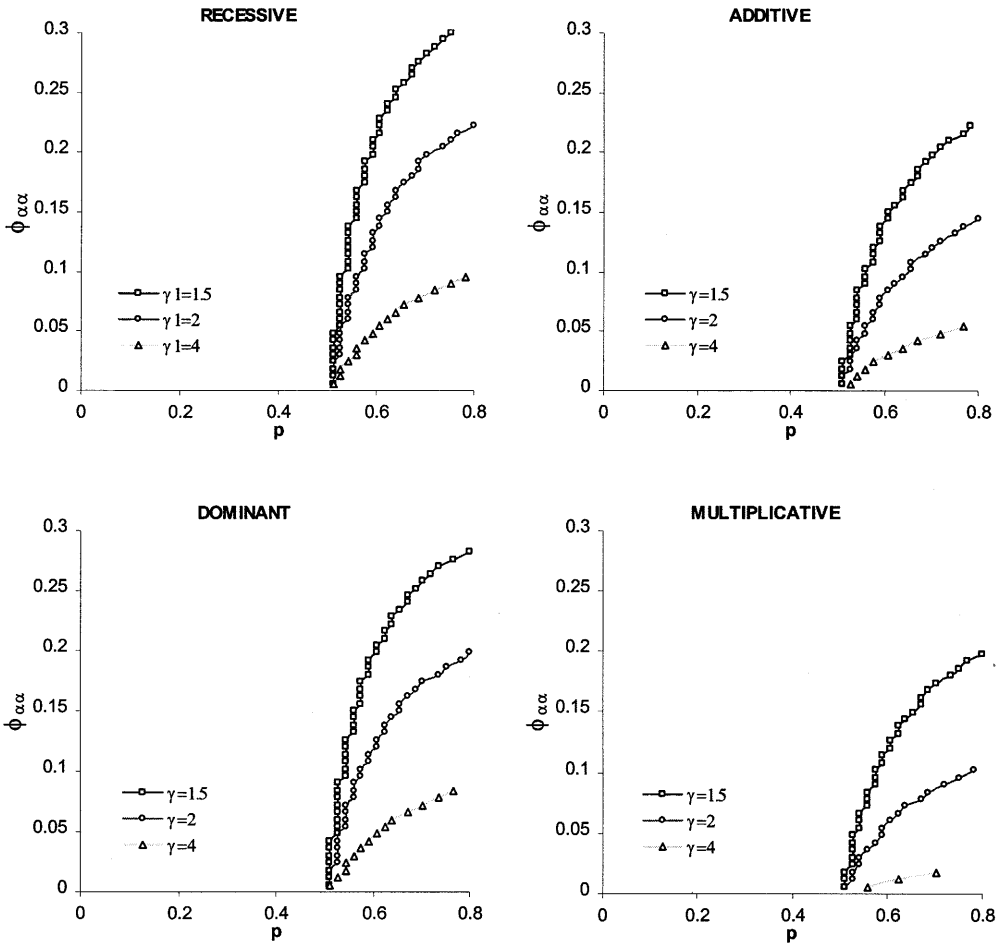

Fig. 3 The regions in the two dimensional parameter (p and ϕ_{aa}) space in which the TDT_{DS} is more powerful than the TDT_{DC} when the marker locus is a DSL per se and $p=f$. In the comparison, η =90% and α= 10^{-7}; again, the choice of η and α is indeed not important for the purpose of comparison of these tests. The *threshold lines* divide the parameter space into two parts. In the parameter space to the *upper left* of the threshold lines, the sample size (N) needed is smaller in the TDT_{DS} than in the TDT_{DC}; thus, the TDT_{DS} is more powerful than the TDT_{DC} in this parameter region. In the parameter space to the *lower right* of the threshold lines, it is the other way around. In the recessive model, γ_1 is specified, and the penetrance for the genotypes AA, Aa, and aa are, respectively, $\gamma_1\phi_{aa}$, ϕ_{aa}, and ϕ_{aa}. For the other three models, γ is specified and γ_1 can be easily inferred from the γ and the genetic models under study

with larger γ_1 or γ, or larger ϕ_{aa}, or larger p, the TDT_{DS} is more likely to be more powerful than the TDT_D. It should be noted that the larger the γ_1 or γ, and/or the larger the ϕ_{aa}, and/or the larger the p, the larger the disease population prevalence ϕ. Even for low p values, the TDT_{DS} can often be more powerful than the TDT_D. Finally, the genetic model (dominant, recessive, etc.) at the DSL is important in determining the parameter space in which the TDT_{DS} is more powerful than the TDT_D. This is apparent when comparing the four plots in Fig. 1 for the same parameters of γ_1 or γ employed in the investigation.

When the marker locus is not a DSL, the conclusions summarized above for the situations when the marker is a DSL per se also hold. In Fig. 2, in the parameter space to the bottom of the threshold line, the TDT_D is more powerful than the TDT_{DS}. On the other side (the upper side) of the threshold line, it is the other way around. In addition, the marker allele frequency f, the recombination rate θ, and the degree of LD (δ) between the marker locus and the DSL all have little effect on the relative power of the TDT_{DS} and TDT_D, although these parameters affect the absolute values of the power of the TDT_{DS} and TDT_D. This is demonstrated by using the recessive model, e.g., as in Fig. 2.

The TDT_{DS} is more powerful than the TDT_{DC} in the majority of the parameter space (Fig. 3). For a frequency (p) of the disease allele A smaller than 0.5, which is probably true in almost all situations, the TDT_{DS} is always more powerful than the TDT_{DC}. Only when ϕ_{aa} is relatively small and p>0.5, may the TDT_{DC} be more powerful than the TDT_{DS}.

The above analyses concentrate on parameter regions in which the TDT_{DS} is more powerful than the TDT_D and the TDT_{DC}. In addition to the limited data in Table 2 and to give a more quantitative indication of the relative power of the TDT_{DS} and the TDT_D, we present, in Fig. 4, the ratio of the sample size needed for the TDT_{DS} and the TDT_D (applied to randomly ascertained family trios each with an affected child) under various parameter values and genetic models. It can be seen (Fig. 4) that the required sample size for 90% power can sometimes be much smaller for the TDT_{DS} than the TDT_D.

Discussion

By utilizing population LD, the TDT has been proposed and employed to identify genes underlying complex traits, while being immune to the problem of population admixture. The customary TDT_D has been applied to affected children and their parents, and most often to case-parent trios. Information from unaffected control-parent trios can

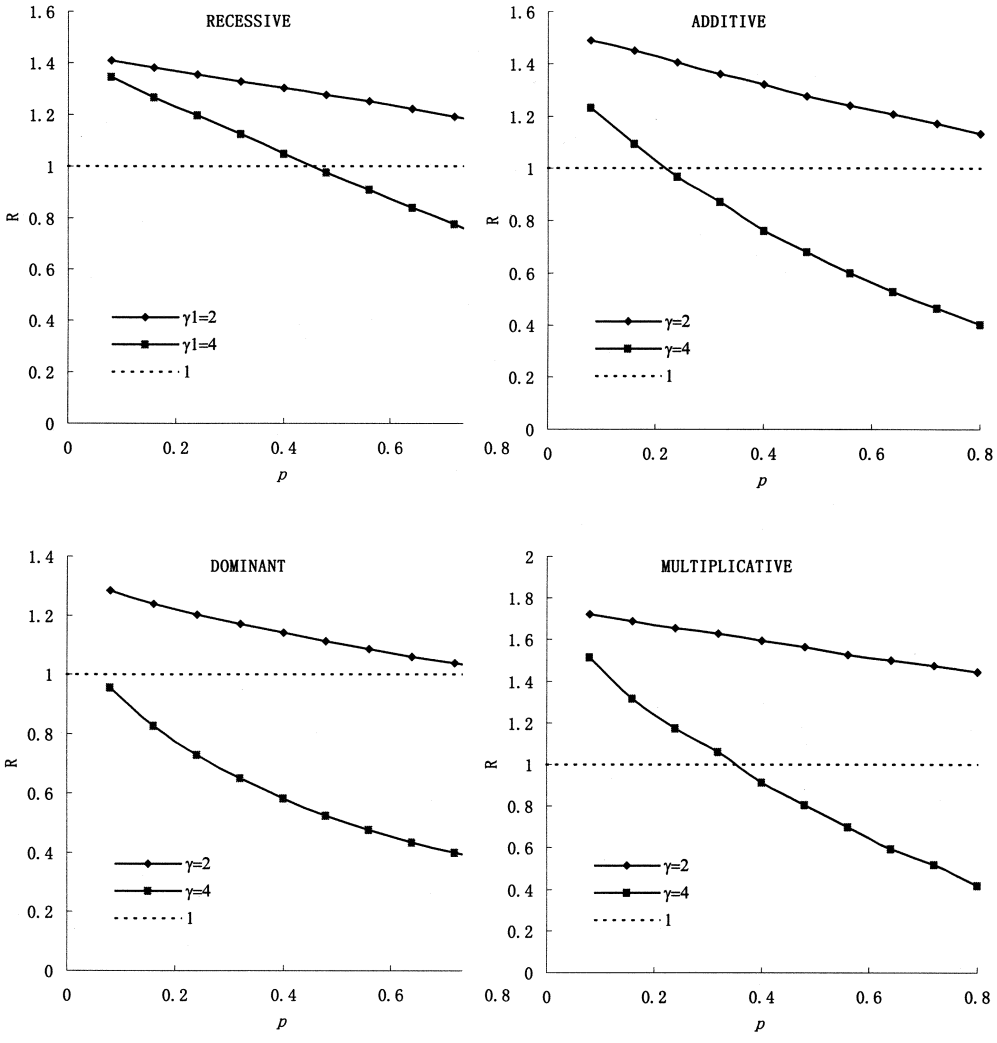

Fig. 4 Numerical comparison of the relative sample sizes needed for the TDT_{DS} and TDT_D in order to reach 90% power. The *X-axis* is the allele frequency *p*. The *Y-axis* is the ratio (*R*) of the number *N* of nuclear family tetrads needed by the TDT_{DS} to that of nuclear family trios needed by the TDT_D. The marker locus is a DSL per se and $p=f$, $\alpha=0.01$, $\phi_{aa}=0.15$ in the recessive and dominant model, $\phi_{aa}=0.1$ in the additive model, and $\phi_{aa}=0.05$ in the multiplicative model

be employed in combination with unrelated case-parent trios for an extension of the TDT (TDT_{DC}) that can be much more powerful than the TDT_D in identifying DSL (Deng and Chen 2001). For many diseases, it is generally easy to obtain unaffected sibs for the affected individuals (Boehnke and Langefeld 1998). Extensions of the TDT have been developed for discordant sibs when the parents are not available (Horvath and Laird 1998; Spielman and Ewens 1998; Boehnke and Langefeld 1998). Whereas these extensions for discordant sibs are valuable for late-onset diseases when parental samples are not available, the power of these TDT extensions is low when compared with the customary TDT_D (Whittaker and Lewis 1999). For complex diseases when parents are available, we present, by utilizing discordant sib-pair-parent tetrads, an extension of the TDT_D, viz., the TDT_{DS}, which can be much more powerful than the TDT_D in a large range of parameter space. The TDT_{DS} is also always more powerful than the TDT_{DC}. We have derived an analytical method for computation of the power of the TDT_{DS} and the power of the TDT_D (applied to the case-parent trios ascertained from the discordant sib-pair-parent tetrads). Our analytical power computation is general in that any genetic model can be easily accounted for. The pedigree disequilibrium test (PDT; Martin et al. 2000, 2001) developed for pedigree analyses may also be applied to the family tetrads with discordant sib pairs. However, our computer simulations have shown (Table 2) that the power of the TDT_{DS} is higher than that of the PDT when the population disease prevalence is high (e.g., >~30%). Hence, whereas the PDT is extremely valuable for TDT analyses of pedigrees that have previously been collected and genotyped, the TDT_{DS} and its investigation here should be useful for designing efficient sampling schemes for family tetrads when the disease is prevalent in a study population.

Segregation distortion is a legitimate concern for significant results obtained in the TDT_D (Spielman et al. 1993; Schaid 1998). With segregation distortion, one allele will be preferentially transmitted to children regardless of the affected status of children. Therefore, unaffected children are necessary in order to rule out the possibility of segregation distortion in yielding significant results in the TDT_D. This is essential in order to validate the significance of a locus tested by the TDT_D in relation to an important DSL. Similar to the TDT_D, the extensions of the TDT tests that have been developed for discordant sib

pairs (without parents, Horvath and Laird 1998; Spielman and Ewens 1998; Boehnke and Langefeld 1998) also may not test for segregation distortion. This is because these extensions test the difference of allele or genotype frequencies in discordant sib pairs and cannot test for the segregation distortion without parental genotype data. In contrast, the TDT_{DC} applied to case-parent trios and unrelated control-parent trios can be employed to test for segregation distortion (Spielman et al. 1993; Deng and Chen 2001). Similarly, the TDT_{DS} applied to discordant sib-pair-parent tetrads can also be employed to test for segregation distortion by the same principle. In the TDT_{DC} and/or the TDT_{DS}, if significant test results are found, and if the same allele is preferentially transmitted to affected and unrelated unaffected children (in the TDT_{DC}) or to the affected and nonaffected sibs (in the TDT_{DS}), then segregation distortion, rather than a significant DSL, is suggested. Otherwise, if significant results are found, and if different alleles are preferentially transmitted to affected and unaffected children (whether they are unrelated or related sibs), segregation distortion can be ruled out, and a significant DSL is suggested. Therefore, both the TDT_{DC} and the TDT_{DS} are significant and useful. First, they can test for segregation distortion that is necessary in order to validate a DSL that is suggested by the TDT_D test. Second, they may dramatically increase the power in detecting a DSL relative to that of the TDT_D, as demonstrated in our study here (for the TDT_{DS}) and in that in Deng and Chen (2001, for the TDT_{DC}).

Compared with the TDT_{DC}, the TDT_{DS} has the following two significant advantages. First, the TDT_{DS} is generally more powerful than the TDT_{DC} and almost always so in biologically plausible situations when the disease allele frequency (p) is less than 0.5. Second, for the same number of affected and nonaffected children employed in analysis, the genotyping for the TDT_{DS} is much reduced, and the ratio of the genotyping effort of the TDT_{DS} to that of the TDT_{DC} is 2:3. This is simply because, for each pair of affected and nonaffected sibs, only a pair of parents needs to be genotyped in the TDT_{DS}, whereas for each pair of affected and nonaffected children, two pairs of parents need to be genotyped in the TDT_{DC}.

Although there is no doubt regarding the necessity of testing for segregation distortion in order to validate a positive result in the TDT_D, the usefulness of unaffected controls in the mapping of DSL is relatively controversial (Schaid 1998; Scott et al. 1999). Our results here and those in Deng and Chen (2001) unambiguously demonstrate that the tests (the TDT_{DS} and TDT_{DC}) that effectively combine controls in analyses can be more powerful for DSL mapping than the TDT_D. We note that, in almost all the current extensions of the TDT analyses, data from only a single type of nuclear family (case-parent trios, or discordant sibs, or discordant sib-pair-parent tetrads, or case-parent and unrelated control trios) are usually employed. However, in practice, we may have mixed samples of these different types of families. The ways in which we can effectively combine the data from these different types of families into a single analysis and investigate the statistical properties (power and size) of the tests for combined data pose a challenge. Permutation-based approaches (e.g., Spielman and Ewens 1998; Boehnke and Langefeld 1998; Deng et al. 2001b) may be suitable for the analyses of combined families of different types. Furthermore, it has been suggested that nuclear families specially ascertained by the affected status of some family members may increase the power of the TDT analyses (Whittaker and Lewis 1998). The combination of data of families randomly ascertained and those ascertained through various schemes and the analytical computation of the associated power also present new challenges. Some specific situations of these challenges have been addressed by authors such as Spielman and Ewens (1998), Knapp (1999), and Martin et al. (2000). Extensions of the TDT to various situations are useful for improving its power and practical significance.

Acknowledgements The investigators of this study were partially supported by grants from Health Future Foundation, NIH K01 grant AR02170–01, NIH R01 grants AR45349-01 and GM60402–01A1, NIH grant P01 DC01813–07, grants from State of Nebraska Cancer and Smoking Related Disease Program (LB598) and Nebraska Tobacco Settlement Fund (LB692), US Department of Energy grant DE-FG03–00ER63000/A00, grants (30025025 and 30170504) from National Science Foundation of China, and grants from the HuNan Normal University and the Ministry of Education of China. W.-M. Chen received a tuition waiver from the Graduate School of Creighton University when conducting the research for this project. We are grateful to the two anonymous reviewers for their generous help and careful comments that improved our manuscript.

References

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

Chakraborty R, Smouse P (1988) Recombination in haplotypes leads to biased estimates of admixture proportions in human populations. Proc Natl Acad Sci USA 85:3071–3074

Chen W-M, Deng H-W (2001) An accurate and general approach for computing the power of the transmission disequilibrium test. Genetic Epidemiol 21:53–67

Deng H-W (2001) Population admixture may change, mask and reverse true genetic effects at genes for complex traits. Genetics 159:1319–1323

Deng H-W, Chen W-M (1999) Re: "Biased tests of association: comparison of allele frequencies when departing from Hardy-Weinberg proportions". Am J Epidemiology 151:335–357

Deng H-W, Chen W-M (2001) The power of the transmission disequilibrium test with both case- and control-parent trios. Genet Res 78:289–302

Deng H-W, Chen W-M, Recker RR (2000) 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 66:1027–1045

Deng H-W, Chen W-M, Recker RR (2001a) Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. Genetics 157:885–897

Deng H-W, Chen W-M, Recker RR (2001b) Powerful and robust tests of linkage and association based on parental genotypes to map loci underlying complex diseases. Life Science Research (in press)

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

Feder JN, Gnirke A, Thomas W, Tsuchihasi Z (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399–408

Hastbacka J, Chapelle A de la, Kaitila I, Sistonen P, Weaver A, Lander ES (1992) Linkage disequilibrium mapping in isolated founder populations: distrophic dysplasia in Finland. Nat Genet 2:204–211

Hastbacka J, Chapelle A de la, Mahtani MM, Clines G, ReeveDaly MP, Daly M, Hamilton BA, et al (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: position cloning by fine-structure linkage disequilibrium mapping. Cell 78:1073–1087

Horvath S, Laird NM (1998) A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 63:1886–1897

Kaplan NL, Martin ER, Weir BS (1997) Power studies for the transmission/disequilibrium tests with multiple alleles. Am J Hum Genet 60:691–702

Knapp M (1999) The transmission disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission disequilibrium test. Am J Hum Genet 64:861–870

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Lazzeroni LC, Lange K (1998) A conditional inference framework for extending the transmission/disequilibrium test. Hum Hered 48:67–81

Martin ER, Monks SA, Warren LL, Kaplan NL (2000) A test for linkage and assoication in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146–154

Martin ER, Bass MP, Kaplan NL (2001) Correcting for potential bias in the pedigree disequilibrium test. Am J Hum Genet 68:1065–1067

Park SK, Miller KW (1988) Random number generator. Commun Assoc Comput Machinery 31:1192–1201

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Schaid DJ (1996) General score tests for asscoiations of genetic markers with diseases using cases and their parents. Genet Epidemiol 13:423–449

Schaid DJ (1998) Transmission disequilibrium, family controls, and great expectations. Am J Hum Genet 63:935–941

Scott LJ, Krolewski A, Rogus JJ (1999) Comparison of the power of the transmission distortion test (TDT) for affected and unaffected trios when disease is highly prevalent. Am J Hum Genet Suppl 65:A397

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

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983–989

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

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506–516

Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland, Mass.

Whittaker JC, Lewis CM (1998) The effects of family structure on linkage tests using allelic association. Am J Hum Genet 63: 889–897

Whittaker JC, Lewis CM (1999) Power comparison of the transmission disequilibrium test and sib-transmission disequilibrium-test statistics. Am J Hum Genet 65:578–580

Xiong MM, Guo SW (1997) Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am J Hum Genet 60:1513–1531