RESEARCH ARTICLE

Transmission/Disequilibrium Tests Incorporating Unaffected Offspring

Qinyu Wei[1], Yuanli Chen[1], Zheng Zeng[1], Chang Shu[1], Lu Long[1], Jianhua Lu[1], Yangxin Huang[2], Ping Yin[1]*

1. Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, **2.** Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, Florida, United States of America

*pingyin2000@126.com

๑ These authors contributed equally to this work.

Abstract

We propose a new method for family-based tests of association and linkage called transmission/disequilibrium tests incorporating unaffected offspring (TDTU). This new approach, constructed based on transmission/disequilibrium tests for quantitative traits (QTDT), provides a natural extension of the transmission/disequilibrium test (TDT) to utilize transmission information from heterozygous parents to their unaffected offspring as well as the affected offspring from ascertained nuclear families. TDTU can be used in various study designs and can accommodate all types of independent nuclear families with at least one affected offspring. When the study sample contains only case-parent trios, the TDTU is equivalent to TDT. Informative-transmission disequilibrium test (i-TDT) and generalized disequilibrium test(GDT) are another two methods that can use information of both unaffected offspring and affected offspring. In contract to i-TDT and GDT, the test statistic of TDTU is simpler and more explicit, and can be implemented more easily. Through computer simulations, we demonstrate that power of the TDTU is slightly higher compared to i-TDT and GDT. All the three methods are more powerful than method that uses affected offspring only, suggesting that unaffected siblings also provide information about linkage and association.

Introduction

Spielman et al. (1993) popularized the transmission/disequilibrium test (TDT) for linkage and association between the marker loci and disease loci for use in studies

of families with at least one affected offspring and two parents [1]. The TDT is a useful method to locate disease genes associated with complex diseases and has the desired property without giving spurious significant results even if there is population stratification which can lead to associations in the absence of linkage (Ewens and Spielman, 1995) [2]. Many extensions and generalizations of the TDT have been developed. Martin et al. (1997) provided a test statistic, TDT_{sp}, that employs the information on transmissions to both members of an affected sib pair and that is valid as a test of both linkage and association [3]. Sun et al. (1999) extended TDT to 1-TDT for use in families with only one of parents available (incomplete nuclear families) [4]. Since the TDT is served as a test for unequal transmission of alleles from the parents to affected offspring, it cannot be performed if the genotypic data for the parents are not available. Horvath and Laird (1998) introduced a discordant-sibship test, the sibship disequilibrium test (SDT), that is used to test data from all of the affected and unaffected siblings [5].

TDT and 1-TDT have the common feature where the offsprings in the data to be analyzed are affected. For diseases with low prevalence, however, it was rather difficult to collect the data. Chao-yu Guo et al. (2007) introduced a new strategy called the informative-transmission disequilibrium test (i-TDT), which uses transmission information from hererozygous parents to all of the affected and unaffected offsprings in ascertained nuclear families and provides a valid chi-square test for both linkage and association [6]. The i-TDT was proved to be more powerful than the Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions (FBAT-o-e) (Lake et al., 2000; Lunetta et al., 2000) which is extended from FBAT (Rabinowitz and Laird, 2000) [7, 8]. Generalized disequilibrium test (GDT) proposed by WeiMin Chen et al. (2009) is another generalization of TDT-like family-based association methods which assesses the genotype difference of all discordant relative pairs in a family and makes use of information beyond first-degree relative pairs [9]. Transmission/disequilibrium tests for quantitative traits (QTDT) proposed by Sun et al. (2000) was used in studies of quantitative trait [10]. The TDT (Spielman et al., 1993) for qualitative traits, the TDT (Rabinovitz, 1997) for quantitative traits and the 1-TDT (Sun et al., 1999) are special cases of QTDT.

This paper is organized as follows. We first introduce our modification on the statistic of QTDT to accommodate studies of qualitative traits and propose a new class of TDT-type tests, transmission/disequilibrium tests incorporating unaffected offspring (TDTU), for use in studies of families with affected and/or unaffected offsprings. The proposed test is valuable and allows researchers full use of the transmission information. Second, we explore the validity and the power of the tests in simulation studies. Finally, we conclude that the new tests are more powerful and allow researchers full use of the available data in detecting linkage between a marker locus and a disease locus.

Methods

General assumptions and notation

Suppose we have $n_i (i=0, 1, 2)$ nuclear families with no, one, and two parents available in the study, respectively, where the subscript 'i' in n_i denotes the number of parents in a family. Let $l_i(j)$ $(i=0, 1, 2; j=1, 2, …, n_i)$ be the number of offsprings in the j-th family among families with i parents. We define the index function $A_{j,k}^{(i)}$ $(k=1,2,…,l_i(j))$ as $A_{j,k}^{(i)} = 1$ if the k-th offspring in the j-th family is affected, $A_{j,k}^{(i)} = 0$ if the k-th offspring in the j-th family is not affected, and $A_{j,k}^{(i)} = -1$ otherwise. We also assume that the marker locus of interest has two alleles 'M' and 'm'. The parents and the offspring are genotyped at a marker locus of interest. For families with both parents, we define the following index function: $Y_{j,k}^{(m)} = 1$ (or -1) if the mother in the j-th family is heterozygous and transmits the M (or m) allele to the k-th offspring, and $Y_{j,k}^{(m)} = 0$ if the mother is homozygous. We similarly define $Y_{j,k}^{(f)}$ for the father. For families with one parent, we only consider offspring-parent pairs with genotypes (mM, mm) or (MM, mM), and offspring-parent pairs with genotypes (mm, mM) or (mM, MM). The first genotype in the bracket is the offspring's genotype and the second genotype is the available parent's genotype (Sun et al, 1999). We define the index function $Y_{j,k}^{(1)}$ as $Y_{j,k}^{(1)} = 1$ if the offspring-parent genotypes are (mM, mm) or (MM, mM), $Y_{j,k}^{(1)} = -1$ if the offspring-parent genotypes are (mm, mM) or (mM, MM), and $Y_{j,k}^{(1)} = 0$ otherwise. Note that the index functions for families with two parents and for families with one parent differ by the superscript.

TDTU when both parents are available

Rabinowitz (1997) first noted that, under the null hypothesis if no linkage between the marker locus and the quantitative loci, the trait value and the index functions $Y_{j,k}^{(f)}$ and $Y_{j,k}^{(m)}$ are conditionally independent, given the parental alleles. In the same way, we can conclude that conditional on the parental alleles, the index function $A_{j,k}^{(2)}$ and the index functions $Y_{j,k}^{(f)}$ and $Y_{j,k}^{(m)}$ are conditionally independent, under the null hypothesis if no linkage between the marker locus and the disease locus. $Y_{j,k}^{(f)}$ and $Y_{j,k}^{(m)}$ are independent under the null hypothesis, that is, $E(Y_{j,k}^{(f)} + Y_{j,k}^{(m)}) = E(Y_{j,k}^{(f)}) + E(Y_{j,k}^{(m)}) = 0$

Thus for any constant c,

$$S_2(c) = \sum_{j=1}^{n_2} \sum_{k=1}^{l_2(j)} (A_{j,k}^{(2)} - c)(Y_{j,k}^{(f)} + Y_{j,k}^{(m)})$$

has mean 0.

The conditional variance of $s_2(c)$ can be estimated by

$$\sigma_2^2(c) = \sum_{j=1}^{n_2} \sum_{k=1}^{l_2(j)} (A_{j,k}^{(2)} - c)^2 (|Y_{j,k}^{(f)}| + |Y_{j,k}^{(m)}|)$$

The class of statistics is given by

$$TDTU = S_2(c)/\sigma_2(c)$$

which has an approximate normal distribution when the number of heterozygous parents is large. When the number of families is small, simulations can be used to determine the p-value.

In this paper, let c be the incidence rate in the population. Let "D" and "d" denote the disease and normal allele. Let f_0, f_1, f_2 denote the probability of being affected when an individual carries 0, 1, 2 risk alleles (the phenocopy rate), and let Aff denote that an individual is diseased. Then c can be calculated by

$$c = E(A) = P(A = 1) \times 1 + P(A = 0) \times 0 = P(A = 1) + P(Aff) \text{ and}$$

$$\begin{aligned} P(Aff) &= P(Aff/DD)P(DD) + P(Aff/Dd)P(Dd) \\ &\quad + P(Aff/dd)P(dd) \\ &= f_2 P(DD) + f_1 P(Dd) + f_0 P(dd) \end{aligned}$$

1-TDTU when only one parent is available

Sun et al. (1999) developed two test statistics for affected offspring applicable to families with one parent. In this paper, we extend those test statistics to affected and unaffected offspring. As in Sun et al. (1999), we assume that the probability of a missing parent is the same for all genotypes at the marker locus of interest. The first class of tests is applicable when either of the following two assumptions, A1 or A2 holds. The second class of tests is applicable even if both assumptions A1 and A2 are violated. However, the second class of tests is, in general, less powerful than the first class of tests.

Assumption A1: Males and females with the same genotype at the marker locus have the same mating preference,

Assumption A2: Both father and mother in each nuclear family are missing with the same probability 1/2 given one of them is missing.

1-TDTU when assumption A1 or assumption A2 holds

Under the null hypothesis of no linkage between the marker locus and the disease locus, $Y_{j,k}^{(1)}$ are independent of $A_{j,k}^{(1)}$. Sun et al. (1999) showed that $P(Y_{j,k}^{(1)} = 1) = P(Y_{j,k}^{(1)} = -1)$ under the null hypothesis of no linkage if assumption A1 or A2 holds. Thus, for any constant c, the conditional mean of

$$S_1(c) = \sum_{j=1}^{n_1} \sum_{k=1}^{l_1(j)} (A_{j,k}^{(1)} - c) Y_{j,k}^{(1)}$$

is zero. Unlike the situation when both parents are available, here $(Y_{j,1}^{(1)}, Y_{j,2}^{(1)}, \ldots, Y_{j,l_1(j)}^{(1)})$ are not independent. The variance of $s_1(c)$ can be estimated by $\sigma_1^2(c) = \sum_{j=1}^{n_1} \left(\sum_{k=1}^{l_1(j)} (A_{j,k}^{(1)} - c) Y_{j,k}^{(1)} \right)^2$. The class of statistical tests for the null hypothesis is given by

$T_1(c) = S_1(c)/\sigma_1(c)$ Note that $T_1(c)$ is the same as the first test statistic in Sun et al. (1999) for any $c \neq 1$.

The 1-TDT when both assumptions A1 and A2 are violated

When both assumptions A1 and A2 are violated, $P(Y_{j,k}^{(1)} = 1) = P(Y_{j,k}^{(1)} = -1)$ no longer holds in general even under the null hypothesis of no linkage between the marker locus and the disease locus. Following Sun et al. (1999), we modify $T_1(c)$ as follows. First, we calculate $s_1(c)$ and $\sigma_1^2(c)$ using families with father available. The corresponding values are denoted by $s_{1f}(c)$ and $\sigma_{1f}^2(c)$, respectively. Similarly, we calculate $s_1(c)$ and $\sigma_1^2(c)$ using families with mother available and the corresponding values are denoted by $s_{1m}(c)$ and $\sigma_{1m}^2(c)$, respectively. Let N_f and N_m be the numbers of families with father and mother available, respectively. Define

$$S_1^*(c) = N_m \times S_{1f}(c) + N_f \times S_{1m}(c),$$

$$\sigma_1^{*2}(c) = N_m^2 \times \sigma_{1f}^2(c) + N_f^2 \times \sigma_{1m}^2(c)$$

The new class of test statistics is then given by

$$T_1^*(c) = S_1^*(c)/\sigma_1^*(c)$$

Simulations

In this section, we evaluate the validity and the power of TDTU, i-TDT, GDT and TDT through simulation studies. All the simulations are accomplished with R2.14.1. We considered additive, recessive, and dominant disease models, and simulated two populations. In the first one, every nuclear family has exactly two offsprings; in the other, every nuclear family has exactly three offsprings. Each population consisted of two sub-populations, and a family belongs to one of them with equal probability 0.5. Under each population, we randomly selected 200 nuclear families with at least one affected offspring under various disease models and different sub-populations. Hence, in the first population, ascertained nuclear

families have either 2 or 1 affected offspring; and in the second population, there would be 3, 2, or 1 affected offspring.

The transmissions from a parent to its affected offspring are correlated if there is linkage, even if there is no association (Spielman and Ewens, 1996). So as a test for linkage TDT is valid for any number of affected offspring in the nuclear families. However, as a test for association either the parental alleles are counted only once or the dependence among offsprings must be allowed in the analysis. Thus, for families with two affected offsprings, we randomly select one to incorporate into analyses for it. TDT ignores information from parents to unaffected offspring. The TDTU, i-TDT and GDT use all transmission information from parents to the offspring regardless of the affection status of the offspring.

We assume that the marker locus of interest has two alleles "M" and "m". Let "D" and "d" denote the disease and normal allele. Then there are 4 possible haplotypes, which can be denoted as MD, Md, mD, md with corresponding relative frequencies p1, p2, p3, p4, respectively. Let f_0, f_1, f_2 denote the probability of being affected when an individual carries 0, 1, 2 risk alleles (the phenocopy rate), respectively. We examined a range of possible values for the disequilibrium coefficient $\delta = p(DM) - p(D)p(M)$. The recombination fraction $\theta = 0.001$. The disease model, phenocopy rate, and the penetrance are indicated in each figure. The additive and recessive results were similar(not shown here).

We simulated 10,000 replicates to check the type-I error at the significance level $\alpha = 0.05$, and 2,000 replicates for the power comparisons. For simulations under the null hypothesis of no linkage or no association, the fraction of times that each test statistic exceeds the critical value is the type-I error. The power of each model is the proportion of significant test under the alternative hypothesis.

Results

We consider three comparisons corresponding to the last three columns in Table 1 in S1 Appendix separately. The first comparison for the four statistics is no association in the presence of linkage; the second one is no linkage in the presence of association, and the last one is no association and no linkage. The TDTU, i-TDT, GDT and TDT all have correct type-I error for testing these three types of composite hypotheses. We can see that, all the three statistics have correct type- I error for testing these three types of composite hypotheses, suggesting that they are valid for both linkage and association.

The simulation results for additive disease models are shown in Figs. 1 and 2. It can be seen that TDT is significantly less powerful than TDTU, i-TDT and GDT due to loss of information. In addition, TDTU is consistently more powerful than i-TDT and GDT.

From Figs. 3–5, we can draw the same conclusion with Figs. 1–2. We can also see that the powers of TDTU, i-TDT and GDT in Figs. 3–5 are higher than that in Figs. 1–2 because of the difference of sample size caused by the offspring number

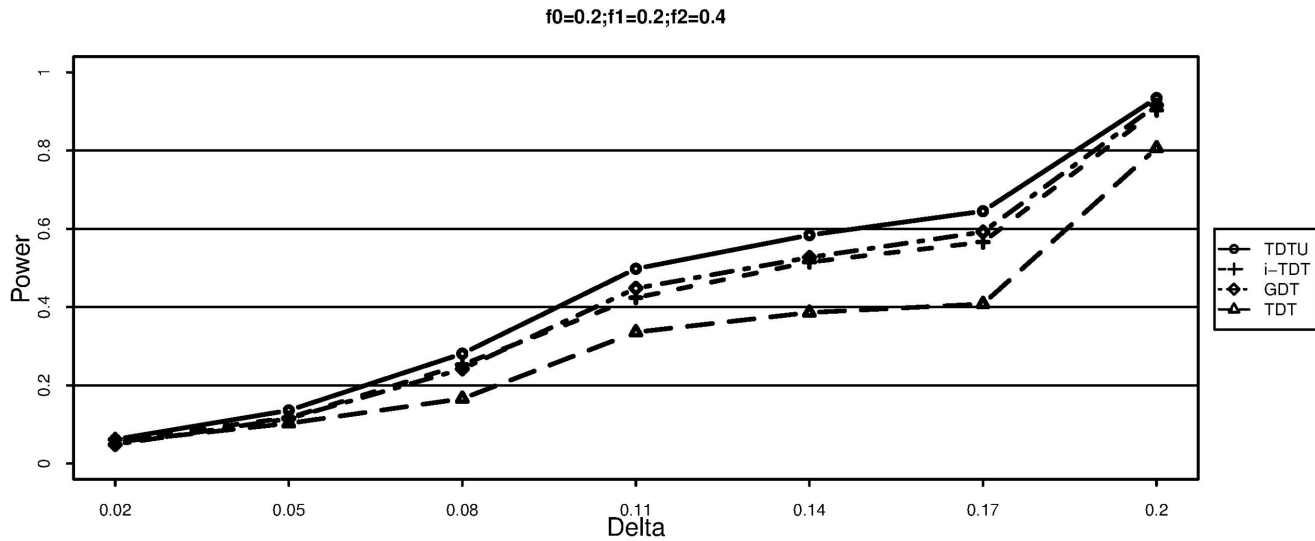**f0=0.2;f1=0.2;f2=0.4**

Fig. 1. Power (each family has two offsprings).

in each family. TDT doesn't get much change since we randomly selected one affected offspring from each family for it. The additive and recessive results are similar (not shown here); more details about the results from simulation studies can be found from tables 1–8 in the S1 Appendix.

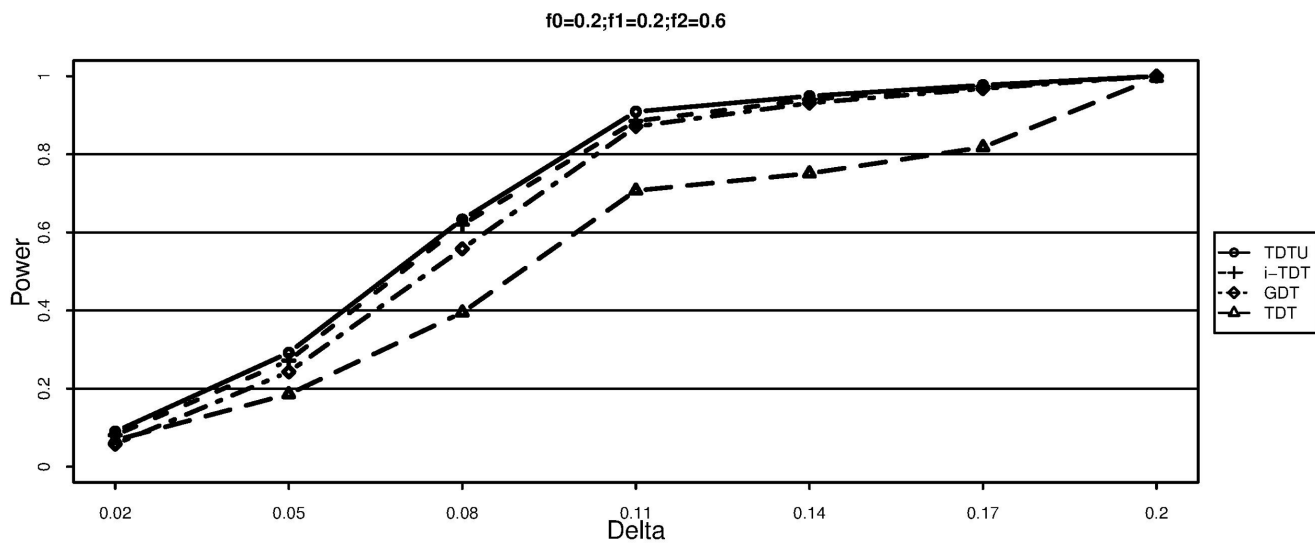**f0=0.2;f1=0.2;f2=0.6**

Fig. 2. Power (each family has two offsprings).

f0=0.2;f1=0.2;f2=0.3

Fig. 3. Power (each family has three offsprings).

Concluding Discussion

In this paper, we first propose a new method for family-based tests of association and linkage called transmission/disequilibrium tests incorporating unaffected offspring (TDTU), which is constructed based on transmission/disequilibrium tests for quantitative traits (QTDT). TDTU can utilize transmission information from heterozygous parents to both affected and unaffected offspring regardless of sibship size and affection status. TDTU can be used in various study designs and
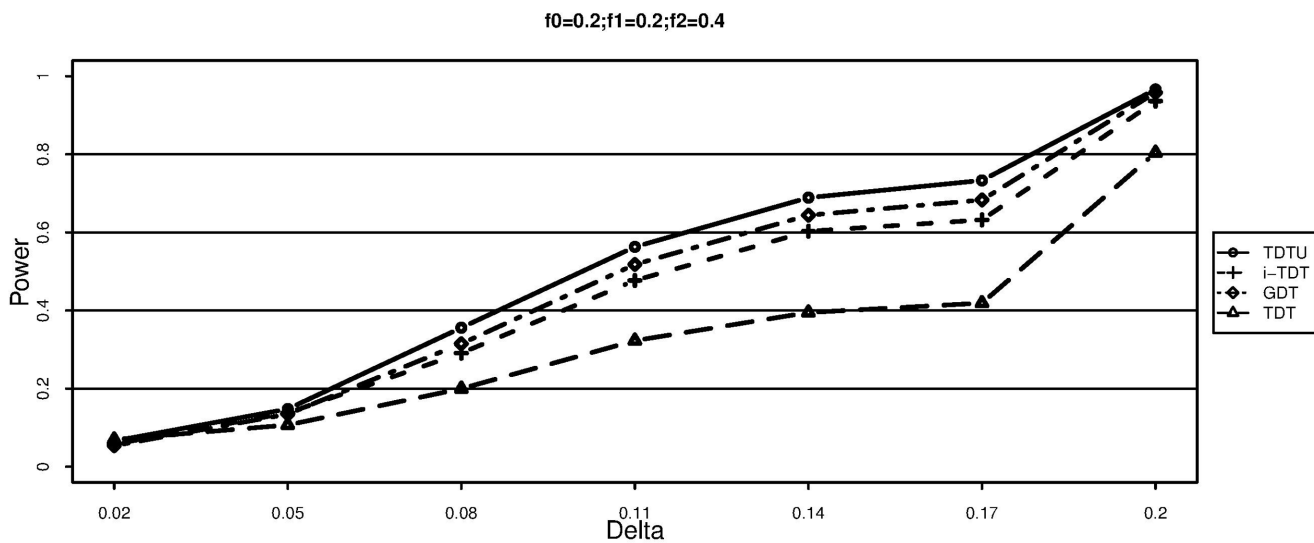
f0=0.2;f1=0.2;f2=0.4

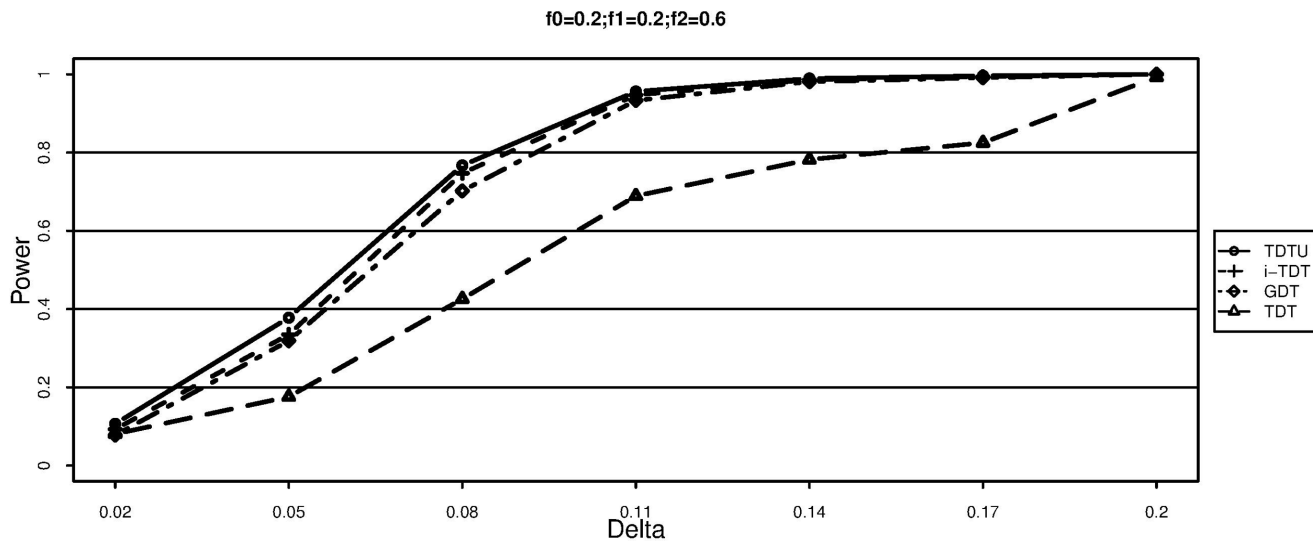**Fig. 4. Power (each family has three offsprings).**

Fig. 5. Power (each family has three offsprings).

doi:10.1371/journal.pone.0114892.g005

can accommodate all types of independent nuclear families with at least one affected offspring. When the study sample contains only case-parent trios, the TDTU is equivalent to TDT. In contract to i-TDT, another method that uses information of both unaffected offspring and affected offspring, the test statistic of TDTU is simpler and more explicit, and can be implemented more easily. From the simulation results above, we can see that power of the TDTU is slightly higher compared to i-TDT and GDT. All TDTU, i-TDT and GDT are more powerful than TDT which uses affected offspring only, suggesting that unaffected siblings also provide information about linkage and association.

Supporting Information

S1 Appendix.
doi:10.1371/journal.pone.0114892.s001 (DOCX)

Acknowledgments

The authors thank Dr Jiyuan Zhou for his assistance with methodological guidance.

Author Contributions

Conceived and designed the experiments: QW. Contributed reagents/materials/analysis tools: ZZ CS LL JL. Wrote the paper: YC. Helped to check the manuscript: YH. Reviewed the calculation and simulation: PY.

References

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

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

3. **Martin ER, Kaplan NL**, **Weir** (1997) Tests for Linkage and Association in Nuclear Families. Am J Hum Genet 61: 439–448.

4. **Sun FZ, Flanders WD, Yang QH, Khoury MJ** (1999) Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. Am J Epidemiol 150: 97–104.

5. **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.

6. **Guo CY, Lunetta KL, DeStefano AL, Ordovas JM, Cupples LA** (2007) Informative-Transmission Disequilibrium Test (i-TDT): combined linkage and association mapping that includes unaffected offspring as well as affected offspring. Genet Epidemiol 31: 115–133.

7. **Lake SL, Blacker D, Laird NM** (2000) Family-Based Tests of Association in the Presence of Linkage. Am J Hum Genet 67: 1515–1525.

8. **Lunetta KL, Faraone SV, Biederman J, Laird NM** (2000) Family-Based Tests of Association and Linkage That Use Unaffected Sibs, Covariates, and Interactions. Am J Hum Genet 66: 605–614.

9. **Chen WM, Manichaikul A, Rich SS** (2009) A Generalized Family-Based Association Test for Dichotomous Traits. Am J Hum Genet 85: 364–376.

10. **Sun FZ, Flanders WD, Yang QH, Zhao HY** (2000) Transmission/disequilibrium tests for quantitative traits. Ann Hum Genet 64: 555–565.