The Transmission Disequilibrium Test and Imprinting Effects Test Based on Case-Parent Pairs

Yue-Qing Hu,1,2 Ji-Yuan Zhou,1,3 Fengzhu Sun,4 and Wing K. Fung1*

1Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China
2Department of Mathematics, Southeast University, Nanjing, China
3Department of Mathematics, Huaihua University, Hunan, China
4Molecular and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles

The transmission disequilibrium test (TDT) based on case-parents trios is a powerful tool in linkage analysis and association studies. When only one parent is available, the 1-TDT is applicable in the absence of imprinting. In the presence of imprinting, a statistic is proposed, based on case-mother pairs and case-father pairs to test for linkage when association is present as well as to test for association when linkage is present. The recombination fractions are allowed to be sex-specific in this test statistic. Meanwhile, a statistic based on case-parent pairs is proposed to test for imprinting. Both test statistics can be extended to include families with more than one affected offspring. A number of simulation studies are conducted to investigate the validity of the proposed tests. The effects of different ratios of the numbers of case-mother pairs and case-father pairs on the powers of the proposed tests are studied through simulation. The results show that the optimal ratio is 1:1. How to combine case-parents, case-mother pairs, and case-father pairs jointly in testing for linkage, association, and imprinting is addressed. Genet. Epidemiol. 31:273–287, 2007. © 2007 Wiley-Liss, Inc.

Key words: association study; case-mother/father; genomic imprinting; genotypic relative risk; linkage analysis; linkage disequilibrium; population stratification; sex-specific recombination fraction; transmission disequilibrium test (TDT)

INTRODUCTION

The transmission disequilibrium test (TDT) based on case-parents trios, introduced originally by Spielman et al. [1993], is a powerful approach to search for genes underlying human complex/common diseases. It tests directly for linkage between the marker locus and a disease susceptibility locus (DSL), when association due to linkage disequilibrium (LD) is present. Although designed as a test for linkage, the TDT is also valid as a test of association in simplex families, even if population structure is present [Spielman and Ewens, 1996]. The TDT essentially tests for the equality of the expected numbers of transmissions and nontransmissions of a marker allele of interest from heterozygous parents to their affected offspring. The TDT requires marker genotypes of affected individuals and their parents. When only one of the parents was available, Sun et al. [1999] proposed the 1-TDT to detect linkage/association between the marker locus and a DSL using genotypes of the affected individual and his/her parent. When both parents’ marker genotypes were unavailable, Spielman and Ewens [1998] proposed the S-TDT for use in sibship with at least one affected individual and one unaffected sibling.

Genomic imprinting, also known as “parent-of-origin effects,” is an important epigenetic factor. There are more and more genes found to be imprinted. Morison et al. [2001] had constructed an imprinted-gene database which contained 488 records at the time of submission (http://igc.otago.ac.nz). For example, parent-of-origin effects have been demonstrated in Beckwith-Wiedemann, Prader-Willi, and Angleman syndromes [Falls et al., 1999]. In genetic studies of case-parents trios,
Weinberg et al. [1998] established a versatile log-linear model for candidate gene to test/estimate LD, maternal effects, and parent-of-origin effects. In testing for LD, the likelihood ratio test outperformed the TDT if the proper genetic model was dominant or recessive, and the reverse was true if a gene–dose effect was the proper model. In the absence of imprinting effects, Weinberg [1999a] considered testing for LD based on the log-linear model allowing for missing parents by employing the EM algorithm. Weinberg [1999b] considered testing for imprinting using the log-linear model with more parsimonious parameters, based on case-parents trios. As pointed out in Weinberg et al. [1998] and Weinberg [1999a,b], their methods were applied to a disease gene. If it is the marker instead of the disease gene itself that is under study, the recombination fraction between the marker locus and the disease gene locus would have to be taken into account in the analysis.

The recombination fractions between the marker locus and a DSL in the meiosis of females and males are often sex-specific. The recombination fraction for human females is on the average 60% higher than that for human males [Fann and Ott, 1995; Broman et al., 1998]. In linkage analysis, imprinting is confounded with differences in recombination fractions for two sexes, and Smalley [1993] suggested the utilization of this information for possible identification of traits undergoing imprinting. The parental-asymmetric test (PAT) Weinberg [1999b] proposed to test for imprinting is applicable to equal recombination fractions for males and females. Recently, Zhou et al. [2006] proposed the parent-of-origin effects test (POET) to test for imprinting based on the marker genotypes of case-parents trios, allowing for sex-specific recombination fractions.

It is likely that researchers obtain genotyping information from families with both parents and families with only one parent. When both parents of the affected child are available, the TDT [Spielman et al., 1993] is applicable to imprinted genes. When only one parent is available, however, we shall show in the later section that the 1-TDT is not applicable to imprinted genes. Hence in this paper we will construct a statistic based on case-mother pairs and case-father pairs to test for linkage when association is present, as well as to test for association when linkage is present for imprinted genes. Meanwhile, a statistic based on families with only one parent is proposed to detect parent-of-origin effects. We also address how to combine case-parents trios, case-mother pairs, and case-father pairs jointly to test for linkage, association or imprinting. The validity of the proposed test statistics is checked through simulation. The effects of different ratios of the numbers of case-mother pairs and case-father pairs on the powers of the proposed tests are investigated. The optimal ratio is found to be 1:1. The tests are also extended to deal with the situation that the parent has more than one affected child.

METHODS

BACKGROUND

Suppose \(D\) and \(d\) are the mutant and normal alleles with population frequencies \(p\) and \(q = 1 - p\) at a DSL, and \(M_1\) and \(M_2\) are the two alleles with population frequencies \(g\) and \(g = 1 - g\) at the diallelic marker locus. The four ordered genotypes at the DSL are \(D/D\), \(D/d\), \(d/D\), and \(d/d\), respectively. The allele before / is paternal and the allele after / is maternal. The four associated risks are denoted by 
\[
\gamma_2 = \frac{\phi_{D/D} \phi_{d/d}}{\phi_{D/d} \phi_{d/d}}, \quad \gamma_1 = \frac{\phi_{D/d} \phi_{d/d}}{\phi_{D/D} \phi_{d/d}}, \quad \gamma_1 = \frac{\phi_{D/D} \phi_{d/d}}{\phi_{D/d} \phi_{d/d}}, \quad \gamma_1 = \frac{\phi_{D/d} \phi_{d/d}}{\phi_{D/D} \phi_{d/d}}.
\]

Let \(\gamma_1 = (\gamma_1 + \gamma_1)/2\) be the average of two heterozygote relative risks. We have \(1 \leq \gamma_1, \gamma_1 \leq 1 \gamma_2\) and \(1 \leq \gamma_1, \gamma_2\). The degree of imprinting is denoted as
\[
I = (\phi_{D/D} - \phi_{d/d})/2
\]
[Strauch et al., 2000]. Thus \(I > 0\) indicates the maternal imprinting or equivalently paternal expression, \(I < 0\) indicates the paternal imprinting or equivalently maternal expression, and \(I = 0\) indicates no imprinting or no effect of the gene on risk. In the case of no imprinting for a disease-related gene, \(\gamma_1 = 1\) means that the mode of inheritance is recessive, \(\gamma_1 = 2\) means dominant, \(\gamma_1 = (1 + \gamma_2)/2\) means additive [Knapp, 1999], and \(\gamma_1 = \sqrt{\gamma_2}\) means multiplicative.

The coefficient of LD is denoted as
\[
\delta = P_{M,D} - gp,
\]
where \(P_{M,D}\) is the haplotype frequency of \(M_1D\). Notice that the frequencies of the four haplotypes can be expressed respectively as
\[
P_{M,D} = gp + \delta, \quad P_{M,d} = gq - \delta, \quad P_{M,d} = gp - \delta, \quad P_{M,d} = gq + \delta.
\]

The marker locus and the DSL are taken to be in LD, i.e., \(\delta \neq 0\), in testing for linkage/imprinting. Notice that the replacement of \(D\) and \(d\), or of \(M_1\) and \(M_2\) will change the sign of \(\delta\) but not its absolute value. Let \(\theta_0\) and \(\theta_m\) be the female and male recombination fractions and
\( \theta = (\theta_f + \theta_m)/2 \) be the sex-average recombination fraction.

It is convenient to use 0, 1, and 2 to represent the marker genotypes \( M_2M_2, M_1M_2, \) and \( M_1M_1, \) respectively. Let \( F, M, \) and \( C \) denote the genotypes of the father, mother, and child, respectively, and so \( F, M, \) and \( C \) take possible values of 0, 1, or 2. Now, we collect \( n_m \) pairs of case-mother each with known marker genotype pair \( MC \) for the mother and affected child, and \( n_f \) pairs of case-father each with known marker genotypes \( FC \) for the father and affected child. It is easy to compare the value of \( M \) and \( C, F \) and \( C \) for each case-parent pair. Let \( N_{M<C} = \sum I_{M<C} \) and \( N_{M>C} = \sum I_{M>C} \) denote the numbers of case-mother pairs in which the mother carries fewer and more copies of marker allele \( M_1 \) than the affected child, respectively, where \( I_{\text{comparison statement}} = 1 \) when the comparison statement holds and is 0 otherwise; let \( N_{F<C} = \sum I_{F<C} \) and \( N_{F>C} = \sum I_{F>C} \) denote the numbers of case-father pairs in which the father carries fewer and more copies of marker allele \( M_1 \) than the affected child, respectively.

We assume throughout this paper that the conditional distribution of the underlying marker genotype trio \( FMC \) given the child is a case corresponding to case-mother pairs and that corresponding to case-father pairs are the same, and the probability of a parent being missing is unrelated to that parent’s genotype. In other words, we assume that there is nondifferential availability or ignorable missingness of parental genotype data. The issue of nonignorable missingness was addressed in Allen et al. [2003] and deserves more attention in the future study. Further, we assume that there are no maternally mediated genetic effects in this study.

In the case of no imprinting and no sex-specific recombination fractions, the 1-TDT [Sun et al., 1999] can be expressed as

\[
1\text{-TDT} = \frac{N_{M<C} - N_{M>C} + N_{F<C} - N_{F>C}}{\sqrt{N_{M<C} + N_{M>C} + N_{F<C} + N_{F>C}}},
\]

where \( N_{M\neq C} = N_{M<C} + N_{M>C} \) and \( N_{F\neq C} = N_{F<C} + N_{F>C} \). In Appendix B, the 1-TDT is shown to be asymptotically normally distributed. When the population is in Hardy-Weinberg equilibrium, the asymptotic distribution of the 1-TDT under the null hypothesis of no LD (i.e., \( \delta(0-0.5) = 0 \)) is \( N(0,1) \). It is also proved in Appendix B that the 1-TDT attains the highest power among the tests in the following class:

\[
T_w = \frac{w(N_{M<C} - N_{M>C}) + (1-w)(N_{F<C} - N_{F>C})}{\sqrt{w^2 N_{M\neq C} + (1-w)^2 N_{F\neq C}}},
\]

\( w \in [0, 1] \).

Note that the 1-TDT is just the \( T_w \) with the particular weight \( w = 0.5 \). It is shown in Appendix B that the \( T_w \) is asymptotically normally distributed.

However, in the case of imprinting, for example, complete paternal imprinting, even under the null hypothesis of no linkage, the mean of the 1-TDT is unknown and could be biased from zero. In fact, it is derived from the results shown in Appendix B, under the null hypothesis of no linkage, that

\[
E(N_{M<C} - N_{M>C}) = n_m \delta I/\phi \quad \text{and} \quad E(N_{F<C} - N_{F>C}) = -n_f \delta I/\phi,
\]

where \( \phi = p^2 \phi_{D/D} + p q \phi_{D/D} + q^2 \phi_{D/D} \) is the population disease prevalence. When \( \delta \neq 0 \), both the expected values \( E(N_{M<C} - N_{M>C}) \) and \( E(N_{F<C} - N_{F>C}) \) are nonzero and further \( E(N_{M<C} - N_{M>C} + N_{F<C} - N_{F>C}) = (n_m - n_f) \delta I/\phi \) is proportional to \( I \), unless \( n_m = n_f \). Also observed is that \( N_{M\neq C} + N_{F\neq C} \) is no longer an unbiased estimator of the variance of \( N_{M<C} - N_{M>C} + N_{F<C} - N_{F>C} \) under the null hypothesis of no linkage, when \( \delta \neq 0 \) with \( n_m \neq n_f \). So the applicability of the 1-TDT as a test of linkage in the presence of association is restricted to the case of no imprinting. Note from the results in Appendix B that the 1-TDT as a test of association in the presence of linkage is applicable to the population in Hardy-Weinberg equilibrium when there are imprinting effects. But this would not be true for the population not in Hardy-Weinberg equilibrium, even in the absence of imprinting effects [Sun et al., 1999]. Simulation results show that the type I error rates of the 1-TDT as a test of linkage in the presence of association as well as a test of association in the presence of linkage could be inflated (see online supplementary tables). Thus, we are confronted with two issues: one is to develop a statistic to test for LD in the presence of imprinting and the other is to detect imprinting effects, based on case-mother and case-father pairs.

In what follows, we turn to seek a suitable weight \( w \) in \( w(N_{M<C} - N_{M>C}) + (1-w)(N_{F<C} - N_{F>C}) \) to construct the required statistic for testing for linkage/association in the presence of imprinting. It is also needed to test if there exist
imprinting effects based on those \( n_m \) case-mother and \( n_p \) case-father pairs.

**TEST FOR LINKAGE/ASSOCIATION**

Taking \( w_0 = n_p/(n_m + n_p) \), it is shown in Appendix C that \( E[w_0(N_{M,<C} - N_{M,>C}) + (1 - w_0)(N_{F,<C} - N_{F,>C})] = 0 \) under the null hypothesis of no linkage (i.e., \( \theta = 0.5 \)). When the Hardy-Weinberg law holds among parents in the source population, it is also shown in Appendix C that \( E[w_0(N_{M,<C} - N_{M,>C}) + (1 - w_0)(N_{F,<C} - N_{F,>C})] = 0 \) under the null hypothesis of no association (i.e., \( \delta = 0 \)). Furthermore, it is verified in Appendix C that \( w_0^2 N_{M,\neq C} + (1 - w_0)^2 N_{F,\neq C} + (n_m + n_p)^{-1}(N_{M,<C} - N_{M,>C})(N_{F,<C} - N_{F,>C}) \) is an unbiased estimator of the variance of \( w_0(N_{M,<C} - N_{M,>C}) + (1 - w_0)(N_{F,<C} - N_{F,>C}) \) under the null hypothesis of no linkage/association. So the 1-TDTI incorporating imprinting can be constructed as follows:

\[
1\text{-TDTI} = \frac{w_0(N_{M,<C} - N_{M,>C}) + (1 - w_0)(N_{F,<C} - N_{F,>C})}{w_0^2 N_{M,\neq C} + (1 - w_0)^2 N_{F,\neq C} + (n_m + n_p)^{-1}(N_{M,<C} - N_{M,>C})(N_{F,<C} - N_{F,>C})}.
\]

Note that the null hypothesis for the 1-TDTI is no linkage or no association, i.e., \( \delta(0-0.5) = 0 \). The statistic 1-TDTI is asymptotically normally distributed. The region of rejection for testing for linkage/association is as follows: \(|1\text{-TDTI}| > z_{\alpha/2}\), where \( z_{\alpha/2} \) is the upper \( \alpha/2 \) point of a standard normal distribution and \( \alpha \) is the significance level. It is to be noticed that, like the TDT and 1-TDT, the 1-TDTI can also be used to test for linkage under association, as well as to test for association under linkage. Both issues are considered in the simulation studies given below.

**TESTING FOR IMPRINTING**

As the 1-TDT [Sun et al., 1999] is not applicable in the presence of imprinting effects, it is desirable to have a test of imprinting. When the population mating is symmetric and the female and male recombination fractions are the same, it is derived in Appendix C that, under the null hypothesis of no imprinting, \( E[w_0(N_{M,<C} - N_{M,>C}) + (1 - w_0)(N_{F,<C} - N_{F,>C})] = 0 \) and \( w_0^2 N_{M,\neq C} + (1 - w_0)^2 N_{F,\neq C} + (n_m + n_p)^{-1}(N_{M,<C} - N_{M,>C})(N_{F,<C} - N_{F,>C}) \) is an unbiased estimator of the variance of \( w_0(N_{M,<C} - N_{M,>C}) + (1 - w_0)(N_{F,<C} - N_{F,>C}) \). So we suggest the POET when only one parent is available as follows:

\[
1\text{-POET} = \frac{w_0(N_{M,<C} - N_{M,>C}) - (1 - w_0)(N_{F,<C} - N_{F,>C})}{w_0^2 N_{M,\neq C} + (1 - w_0)^2 N_{F,\neq C} + (n_m + n_p)^{-1}(N_{M,<C} - N_{M,>C})(N_{F,<C} - N_{F,>C})}.
\]

The statistic follows asymptotically a normal distribution and the region of rejection for testing for imprinting is as follows: \(|1\text{-POET}| > z_{\alpha/2}\).

**PARENT WITH MORE THAN ONE CHILD**

When some mothers/fathers have more than one affected child, the statistics (2) and (3) should be adjusted accordingly to test for LD and imprinting, respectively. In this situation, the term \( N_{M,<C} - N_{M,>C} \) or equivalently \( \Sigma(I_{M,<C} - I_{M,>C}) \) in the numerator of equations (2) and (3) is replaced by \( \Sigma_z \Sigma(I_{M,<C} - I_{M,>C}) \) where the first summation sums over all mothers \( M \) and the second summation is for all possible children \( C \) with the same mother \( M \). Similarly, the term \( N_{F,<C} - N_{F,>C} \) is replaced by \( \Sigma_z \Sigma(I_{F,<C} - I_{F,>C}) \). It is verified in Appendix C that the unbiased estimator of the variance of

\[
T = w_0 \sum_M \sum_C (I_{M,<C} - I_{M,>C})
\]

\[
\pm (1 - w_0) \sum_F \sum_C (I_{F,<C} - I_{F,>C})
\]

under the null hypothesis of no LD/imprinting is

\[
\text{Var}_0(T) = w_0^2 \left( \sum_M \sum_C I_{M,\neq C} + \sum_M \sum_{C,C'} (I_{M,<C} - I_{M,>C}) \times (I_{M,<C} - I_{M,>C}) \right)
\]

\[
+ (1 - w_0)^2 \left( \sum_F \sum_C I_{F,\neq C} + \sum_F \sum_{C,C'} (I_{F,<C} - I_{F,>C}) \times (I_{F,<C} - I_{F,>C}) \right)
\]

\[
\pm w_p^2 \sum_{i,j} n_{m_{i,j}} n_{p_{i,j}} + w_p^2 \sum_{i,j} n_{i,j} \sum_{i,j} n_{i,j} \sum_M \sum_C (I_{M,<C} - I_{M,>C})
\]

\[
\times \sum_F \sum_C (I_{F,<C} - I_{F,>C}),
\]

where \( n_{m_{i,j}} \) is the number of families in which the mother has \( i \) affected children, \( n_{p_{i,j}} \) is the number of families in which the father has \( i \) affected children, \( i = 1, 2, \ldots \), \( n_m = \Sigma_n n_{m_{i,j}} \), \( n_p = \Sigma_n n_{p_{i,j}} \), and the summation \( \Sigma_{i,j} \) sums over all combinations of children \( C_m \) and \( C_p \) with the same parent. So the
required test statistics can be derived accordingly as $T/\sqrt{\text{Var}_0(T)}$.

**SIMULATION RESULTS**

Since a number of parameters are involved, we fix the following parameter values throughout the simulation study for illustration purpose: $\phi_{D/D} = 0.6, \phi_{d/d} = 0.2$, and $\gamma_1$ and $\gamma_1m$ are fixed at some values between 1 and $\gamma_2 = 3$. The other parameter values will be specified later. For each set of parameter values, we run the simulation 20,000 times and the actual sizes/powers are estimated as the proportions of replicates in which the null hypothesis is rejected at significance level $\alpha$ when the simulation is performed under the null/alternative hypothesis. Two significance levels 5 and 0.5% are used to evaluate the sizes

**MODEL FOR TESTING FOR LINKAGE**

The population stratification demographic model [Sun et al., 1999] is used to assess the 1-TDTI as a test of linkage and the 1-POET. The population under study is composed of two subpopulations. The frequencies of haplotypes $M_d$, $M_dD$, $M_dD$, and $M_1D$ in the first (second) population are set to 0.47, 0.03, 0.03, and 0.47 (0.4, 0.1, 0.1, and 0.4), respectively. The first subpopulation is 70% of the total population and the second one is 30%. For simplicity, each of the two subpopulations is assumed to be in Hardy-Weinberg equilibrium, even though the resulting mixed population is not and mating is not random in the mixed population. In each subpopulation, we first generate haplotypes at the marker locus and a DSL for the father and mother according to those four haplotype frequencies. The paternal haplotype of the child is generated from the father’s haplotypes with the male recombination fraction $\theta_m$. Similarly, the maternal haplotype of the child is generated. The affection status of the child is determined by the child’s genotype at the disease locus and the associated four risks $\phi_{D/D}, \phi_{D/d}, \phi_{d/D},$ and $\phi_{d/d}$. The properties of the 1-TDTI as a test of linkage in terms of the powers and type I error rates are explored in a number of situations, including moderate sample sizes and different combinations of the numbers of case-mother pairs and case-father pairs.

**MODEL FOR TESTING FOR ASSOCIATION**

To investigate the properties of the 1-TDTI as a test of association, we adopt the following assortative mating demographic model [Sun et al., 2000] in which Hardy-Weinberg equilibrium again does not hold. In this population, 70% of the families were formed through random mating and 30% of the families were formed through assortative mating where the mother carries more copies of marker allele $M_1$ than the father, both the disease allele frequency and the marker allele frequency are 0.5, and both the female and male recombination fractions are 0.001, while the other parameter values are taken the same as before. The mechanism of data generating is similar to that described above. The properties of the 1-TDTI as a test of association in terms of the powers and type I error rates are explored for various $\gamma_1$ values, imprinting degrees, and different combinations of the numbers of case-mother and case-father pairs.

**MODEL FOR TESTING FOR IMPRINTING**

The population stratification demographic model described above is also used to assess the 1-POET as a test of imprinting. The properties of the 1-POET in terms of the type I error rates are investigated for various $\gamma_1$ values, female and male recombination fractions, and different sample sizes. The properties of the 1-POET in terms of the powers are investigated in the cases of complete paternal imprinting, incomplete paternal imprinting, incomplete maternal imprinting, and complete maternal imprinting.

In the case of the parent having more than one affected child, we choose the following sample sizes for the simulation models given above: for case-mother pairs, there are 90 families with one affected child and 40 families with two affected children; for case-father pairs, there are 70 families with one affected child and 30 families with two affected children.

**SIZES OF THE 1-TDTI AS A TEST OF LINKAGE/ASSOCIATION**

We are going to investigate the actual type I error rates of the test. For the completeness of the investigation, we choose 13 representative pairs of values for $\gamma_1$ and $\gamma_1m$ (which are equivalently expressed as $\gamma_1$ and $J$ in Tables I and II), which are scattered uniformly in the square [Strauch et al., 2000] composed of $\{\gamma_1, \gamma_1m\} | 1 \leq \gamma_1, \gamma_1m \leq \gamma_2$ (see Tables I and II for details). It is noted that
TABLE I. Type I error rates (%) of the 1-TDTI as a test of linkage in the presence of association at significance level $\alpha = 5$ and 0.5% for simulation with 20,000 replicates in the population stratification demographic model having $\theta_f = 0_m = 0.5$ and $\delta = 0.22$

<table>
<thead>
<tr>
<th>$\gamma_1$</th>
<th>$I$</th>
<th>Sample size pair ($n_m, n_p$) and $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(100, 100)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>$1+3\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.78</td>
</tr>
<tr>
<td>$1+3\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.83</td>
</tr>
<tr>
<td>$1+3\gamma_2$</td>
<td>0</td>
<td>4.72</td>
</tr>
<tr>
<td>$1+3\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.89</td>
</tr>
<tr>
<td>$1+3\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>5.12</td>
</tr>
<tr>
<td>$1+3\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.68</td>
</tr>
<tr>
<td>$1+\gamma_2$</td>
<td>0</td>
<td>4.57</td>
</tr>
<tr>
<td>$1+\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.78</td>
</tr>
<tr>
<td>$1+\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.47</td>
</tr>
<tr>
<td>$1+\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.90</td>
</tr>
<tr>
<td>$1+\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.60</td>
</tr>
<tr>
<td>$1+\gamma_2$</td>
<td>$1^{-1} \phi_{d/d}$</td>
<td>4.52</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>4.65</td>
</tr>
</tbody>
</table>

($\gamma_{1p}, \gamma_{1m}$) = ($\gamma_2$, $\gamma_2$) corresponds to the common dominant mode of inheritance ($\gamma_{1p}, \gamma_{1m}$) = ((1 + $\gamma_2$)/2, (1 + $\gamma_2$)/2) corresponds to the additive mode of inheritance [Knapp, 1999], and ($\gamma_{1p}, \gamma_{1m}$) = (1, 1) corresponds to the common recessive mode of inheritance, ($\gamma_{1p}, \gamma_{1m}$) = ($\gamma_2$, 1) indicates complete maternal imprinting, and ($\gamma_{1p}, \gamma_{1m}$) = (1, $\gamma_2$) indicates complete paternal imprinting. Furthermore, we choose the numbers of case-mother and case-father pairs as ($n_m, n_p$) = (100, 100), (100, 200), and (200, 100).

Table I reports the actual sizes of the 1-TDTI as a test of linkage obtained by simulation in the population stratification demographic model, where both the female and male recombination fractions are taken to be 0.5. Table II reports the actual sizes of the 1-TDTI as a test of association in the assortative mating demographic model with $\theta_f = 0_m = 0.001$, where the coefficient of LD is taken to be 0. Most of the entries in Tables I and II show that the sizes of 1-TDTI are close to but marginally lower than the nominal 5 and 0.5% levels, respectively, which indicates a slight conservativeness of the 1-TDTI. In the evaluation of the sizes of the 1-POET, for a given $\gamma_1$, $\gamma_2$ is taken as the following five values which are equally spaced in the range of 1 and $\gamma_2$: 1, (3 + $\gamma_2$)/4, (1 + $\gamma_2$), (1 + 3$\gamma_2$)/4, and $\gamma_2$. Table III shows that the actual sizes are close to but

---

*Hu et al.*

slightly lower than the nominal ones under the situation when $\theta_f = \theta_m = 0.001$, and that when $\theta_f = 0.01$ and $\theta_m = 0.001$, where the degree of imprinting is taken to be 0. Most of the entries show that the 1-POET, like the 1-TDTI, is slightly conservative. Furthermore, we evaluate the sizes when $\theta_f = 0.03, 0.05, 0.1,$ and $0.3$ while $\theta_m = 0.01$ or $0.1$, respectively, and the actual sizes range from 4.68 to 5.16% for the nominal 5% level, and from 0.34 to 0.51% for the nominal 0.5% level. So it is illustrated in our simulation studies that the 1-POET seems still applicable to detect parent-
of-origin effects when the female and male recombination fractions are different.

When the mother/father has more than one affected child, the sizes of the 1-POET are 4.82% when \( \theta_f = \theta_m = 0.001 \) and are 4.88% when \( \theta_f = 0.01 \) and \( \theta_m = 0.001 \), compared with the nominal 5% level. If the nominal level is set with 0.5%, then the corresponding sizes are 0.42 and 0.37%, respectively. This illustrates that we can also use the 1-POET to deal with the case of parent having more than one affected child.

POWERS OF THE 1-TDTI WITH DIFFERENT \( n_m:n_p \)

The objective of this section is to investigate the effects of different sample sizes ratio \( n_m:n_p \) on the performance of the proposed 1-TDTI as a test of linkage/association when the sum \( n_m+n_p \) is fixed. The findings can provide some guidelines in collecting such data in conducting linkage/association analysis. For illustration purpose, we fix \( \gamma_1 \) at \( (1 + \gamma_2)/2 \) and the sum \( n_m+n_p \) at 200, to calculate the powers of the 1-TDTI when the number of case-mother pairs \( n_m \) takes values from 50 to 150 in increments of 10.

First, we simulate the powers of the 1-TDTI as a test of linkage when \( \gamma_f = \gamma_m = 0.001 \) and that as a test of association when \( \delta = 0.22 \) in the cases of (a) \( I = (1 - \gamma_2)\gamma_d/d/2 \) (complete paternal imprinting), (b) \( I = 7(1 - \gamma_2)\gamma_d/d/16 \) (incomplete paternal imprinting), (c) \( I = 7(\gamma_2 - 1)\gamma_d/d/16 \) (incomplete maternal imprinting), (d) \( I = (\gamma_2 - 1)\gamma_d/d/2 \) (complete maternal imprinting). Figure 1 depicts the power of the 1-TDTI as a test of linkage against the number of case-mother pairs \( n_m \) in the population stratification demographic model, and Figure 2 depicts the power of the 1-TDTI as a test of association against \( n_m \) in the assortative mating demographic model. The effects of different sample size ratio \( n_m:n_p \) on the power of the 1-TDTI could be substantial, though the sum \( n_m+n_p \) is the same. The optimal choice of \( n_m:n_p \) is 1:1, which signifies that case-father and case-mother pairs are equally important in testing for LD. Figure 1 shows that the difference among the powers of the 1-TDTI as a test of linkage with complete paternal imprinting, incomplete

![Fig. 1. The actual powers of the 1-TDTI as a test of linkage are plotted against the number of case-mother pairs \( n_m \) in increments of 10 under (a) complete paternal imprinting (\( \gamma_d/d = 0.2 \), \( \gamma_d/D = 0.06 \)); (b) incomplete paternal imprinting (\( \gamma_d/d = 0.175 \), \( \gamma_d/D = 0.575 \)); (c) incomplete maternal imprinting (\( \gamma_d/d = 0.575 \), \( \gamma_d/D = 0.175 \)); (d) complete maternal imprinting (\( \gamma_d/d = 0.6 \), \( \gamma_d/D = 0.2 \), \( \gamma_1 = (1 + \gamma_2)/2 \), \( \delta = 0.001 \), and \( n_m + n_p = 200 \) in the population stratification demographic model. Powers are based on 20,000 replicates and assessed at the 5% level.](genet-epidemiol-10.1002/gepi.280-fig1.png)
Paternal imprinting, incomplete maternal imprinting and complete maternal imprinting is very small. Figure 2 shows that the difference between the powers of the 1-TDTI as a test of association with complete paternal imprinting and incomplete paternal imprinting is very small, that with complete maternal imprinting and incomplete maternal imprinting is also very small, and that with complete paternal imprinting and complete maternal imprinting is about 3%.

POWERS OF THE 1-POET WITH DIFFERENT \( n_m : n_p \)

We evaluate the powers of the 1-POET in the population stratification demographic model in the cases of (a) complete paternal imprinting, (b) incomplete paternal imprinting, (c) incomplete maternal imprinting, and (d) complete maternal imprinting as described in the previous section, where both the female and male recombination fractions are taken as 0.001. Figure 3 plots the corresponding power of the 1-POET against the number of case-mother pairs \( n_m \). It shows that the effects of the sample size ratio \( n_m : n_p \) on the power of the 1-POET could be substantial. The case-mother and case-father pairs are equally important in testing for imprinting using the 1-POET and the optimal ratio of \( n_m : n_p \) is again 1:1. The difference between the powers of the 1-POET in the cases of complete paternal imprinting and complete maternal imprinting is also very small, and that in the cases of complete paternal imprinting and incomplete maternal imprinting is also very small. However, there is about 10% difference between the powers of the 1-POET in the cases of complete imprinting and incomplete imprinting.

DISCUSSION

In principle, we can choose different \( w \) in \( w(N_{M<C} - N_{M>C}) + (1 - w)(N_{F<C} - N_{F>C}) \) to construct a test statistic to test for LD. But if we do so, we have to estimate the mean of the statistic under the null hypothesis of no LD and this estimation is usually difficult to deal with. In view of this,
we choose a suitable weight \( w_0 = \frac{n_p}{n_m + n_p} \) such that the mean of \( w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C}) \) under the null hypothesis is zero and thus we only need to give an unbiased estimator of the variance. From the simulation results, the simulated sizes are somehow a bit lower than the nominal ones, which is probably due to the nonasymptotic behavior when the sample sizes \( n_m \) and \( n_p \) are not large enough. We also have done further simulations with larger sample sizes that have better asymptotic behavior.

In theory, we require \( \theta_f = \theta_m \) to guarantee the expectation of \( w_0(N_{M<C} - N_{M>C}) - (1 - w_0)(N_{F<C} - N_{F>C}) \) being 0 under the null hypothesis of no imprinting. In the presence of association between the marker locus and a DSL, it is plausible to assume that both the female and male recombination fractions are small, say less than 0.01, and the difference between the sex-specific recombination fractions is consequently small. So in practice, the 1-POET could still be applicable. In fact, our simulation studies show that the 1-POET may still be used even when the difference between the recombination fractions is larger than 0.01.

It is noted that the weight \( w_0 = \frac{n_p}{n_m + n_p} \) is employed in constructing the 1-TDTI and 1-POET. Notice that there are \( n_m \) case-mother and \( n_p \) case-father pairs. So in testing for LD/imprinting the weight \( w_0 = \frac{n_p}{n_m + n_p} \) makes the contribution of case-mother pairs the same as that of case-father pairs. It would be expected that a balanced design of case-father and case-mother pairs would provide the most efficient information in testing for LD/imprinting. It is also observed in our simulation studies that the power of the 1-TDTI/1-POET attains the highest value in the case of equal numbers of case-mother and case-father pairs.

In practice, it is common to have two kinds of data, one from families with both parents and the other from families with only one parent. When the parents of the affected child are available, the conventional TDT [Spielman et al., 1993] can be expressed as \( (T - NT)/\sqrt{T + NT} \), where \( T \) and \( NT \) denote the numbers of transmissions and nontransmissions of marker allele

**Fig. 3.** The actual powers of the 1-POET are plotted against the number of case-mother pairs \( n_m \in [50, 150] \) in increments of 10 under (a) complete paternal imprinting \( (\phi_E = 0.2, \phi_D = 0.6) \); (b) incomplete paternal imprinting \( (\phi_E = 0.175, \phi_D = 0.575) \); (c) incomplete maternal imprinting \( (\phi_E = 0.575, \phi_D = 0.175) \); (d) complete maternal imprinting \( (\phi_E = 0.6, \phi_D = 0.2) \), having \( \phi_E = 0.6, \phi_D = 0.2, \gamma_1 = (1 + \gamma)/2, \gamma = \theta_m = 0.001 \), and \( n_m + n_p = 200 \) in the population stratification demographic model. Powers are based on 20,000 replicates and assessed at the 5% level.
where $NF$ tests for the equality of numbers of case-parents proposed to test for imprinting essentially of combining both kinds of data for linkage/affected offspring, respectively. Thus, one way to detect these different effects and take account of genotype effects into analysis. How to differentiate genotype effects and incorporate the maternal log-linear model to estimate/detect the maternal effects, population mating symmetry and equal recombination fractions for males and females. It is noted from the combined test statistic that NF under study is not a DSL, Weinberg’s [1999b] method is a very powerful one when the marker locus under study is a candidate DSL. In this situation, the terms $NF>M$ and $NF<M$ in equation (5) can be replaced respectively by $NF>M,C=1$ and $NF<M,C=1$. When the marker locus under study is not a DSL, Weinberg’s [1999b] method is also applicable under no maternal effects, population mating symmetry and equal recombination fractions for males and females.

In this paper, we consider the detection of parent-of-origin effects and the test for LD for imprinted disease genes. In fact, there have been increasing interests in maternal genotype effects. For example, Weinberg et al. [1998] established a log-linear model to estimate/detect the maternal genotype effects and incorporate the maternal genotype effects into analysis. How to differentiate these different effects and take account of these effects into linkage analysis and association studies deserve future investigation.

**ACKNOWLEDGMENTS**

The authors are grateful to three anonymous reviewers for insightful comments which improved greatly the presentation of the material. This work was partially supported by a Hong Kong RGC Research Grant, the National Natural Science Foundation of China (10329102, 10561008), and the Scientific Research Fund of Huaihua University. F. Sun thanks the support of Dr. Fung during his sabbatical leave in the University of Hong Kong.

**ELECTRONIC DATABASE INFORMATION**

http://www.hku.hk/statistics/staff/wingfung/1TDTI and Imprinting Effects Test

http://www.hku.hk/statistics/staff/wingfung/1TDT/ for downloading the 1-TDTI and the 1-POET software packages.

http://www.hku.hk/statistics/staff/wingfung/1TDTSizes.pdf for supplementary online tables of the sizes of the 1-TDT as a test of linkage in the presence of association as well as a test of association in the presence of linkage.

**REFERENCES**


child is affected

For the

(see the first two columns in Table IV).

case-parents trios are classified into 15 categories

length as vector

For any two constant vectors

r

1983],

r

1999. Transmission
disequilibrium test (TDT) when only one parent is available:

disequilibrium tests for quantitative traits. Ann Hum Genet 64:

555–565.

APPENDIX A

PRELIMINARY

Let \((N_1, \ldots, N_k, n - \sum_{i=1}^{k} N_i)\) be the multinomial

distribution \(M(r_1, \ldots, r_k, 1 - \sum_{i=1}^{k} r_i; n)\) [Lehmann,

1983], \(r = (r_j)_{j=1}^{k}\), and \(N = (N_j)_{j=1}^{k}\). Then \(N/n\) is the

maximum likelihood estimator of the parameter vector \(r\). So we have [Rao, 1973]

\[
\frac{N}{n} \rightarrow r\quad \text{in probability,}
\]

\[
\frac{N - nr}{\sqrt{n}} \rightarrow N(0, \text{diag}(r) - rr^T)\quad \text{in law.}
\]

For any two constant vectors \(u\) and \(v\) of the same

length as vector \(r\), we have

\[
(u + v)^TN/N \rightarrow (u + v)^Tr\quad \text{in probability,}
\]

\[
(u - v)^TN - n(u - v)^Tr/N \rightarrow N(0, (u - v)^T(\text{diag}(r) - rr^T)\times (u - v))\quad \text{in law}
\]

APPENDIX B

ASYMPTOTIC DISTRIBUTION AND

OPTIMALITY: \(T_w\)

Based on the marker genotypes \(FMC\), the

case-parents trios are classified into 15 categories

(refer to the first two columns in Table IV).

For the \(j\)th (\(1 \leq j \leq 15\)) category, let \(s_j\) denote the

conditional probability that a family falls into this category, given the child is a case. For example, \(s_1 = P(F = 2, M = 1, C = 2|\)

child is affected) and \(s_2 = P(F = 1, M = 2,\)

\(C = 2|\) child is affected). Moreover, for each family

in the \(j\)th category (\(1 \leq j \leq 15\)), let \(u_{mj}\) be 1 if the

mother has fewer copies of marker allele \(M_1\) than the

affected child and 0 otherwise, and let \(v_{mj}\) be 1 if the mother has more copies of marker allele \(M_1\) than the affected child and 0 otherwise. Similarly, we define \(u_{pj}\) and \(v_{pj}\) for the father and affected

child. See Table IV for details. Denote

\[
s = (s_j)_{j=1}^{14}, u_m = (u_{mj})_{j=1}^{14}, v_m = (v_{mj})_{j=1}^{14}, u_p = (u_{pj})_{j=1}^{14},\]

and \(v_p = (v_{pj})_{j=1}^{14}\).

Notice that every case-mother pair is deduced from a case-parents trio when the father is missing. For the \(n_m\) case-mother pairs, there are \(n_m\) underlying case-parents trios. Let \(N_m = (N_{mj})_{j=1}^{14}\), where \(N_{mj}\) is the number of families falling into the \(j\)th category (see Table IV for details) among those \(n_m\) case-parents trios,

<table>
<thead>
<tr>
<th>(j)</th>
<th>(FMC)</th>
<th>(u_{mj})</th>
<th>(v_{mj})</th>
<th>(u_{pj})</th>
<th>(v_{pj})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>212</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>122</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>211</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>121</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>112</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>111</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>110</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>101</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>011</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>010</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>222</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>201</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>021</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(u_{mj} = I_{M<C}, v_{mj} = I_{M>C}, u_{pj} = I_{F<C},\)

and \(v_{pj} = I_{F>C}, 1 \leq j \leq 15\).

Hu et al.

Genet. Epidemiol. DOI 10.1002/gepi
Similarly, define $N_p = (N_{p,i})_{i=1}^{14}$ for the $n_p$ case-parent trios. So $(N_{m,1}, \ldots, N_{m,14}, n_m - \sum_{i=1}^{14} n_m,i)$ follows the multinomial distribution $M(s_1, \ldots, s_{14}, 1 - \sum_{i=1}^{14} s_i; n_m)$, $(N_{p,1}, \ldots, N_{p,14}, n_p - \sum_{i=1}^{14} N_{p,i})$ follows the multinomial distribution $M(s_1, \ldots, s_{14}, 1 - \sum_{i=1}^{14} s_i; n_p)$, and $N_m$ and $N_p$ are independent.

Based on $u_m, v_m, u_p, p_v, N_m, N_p$, and $n_p$, we have

$$
\Sigma_{M<i} = u_m^T N_m, \quad \Sigma_{M<i'} = v_m^T N_m, \quad \Sigma_{I_F<i} = u_p^T N_p, \quad \text{and} \quad \Sigma_{I_F<i'} = v_p^T N_p.
$$

Applying equations (6) and (7) to $N_m \sim M(s, n_m)$ and the constant vectors $u_m$ and $v_m$ and observing $(u_m - v_m)^T \text{diag}(s)$

$$
(u_m - v_m) = \sum_{i=1}^{14} (u_m - v_m)^2 s_i = \sum_{i=1}^{14} (u_m + v_m)^2 s_i = (u_m + v_m)^T s,
$$

we have

$$
\frac{(u_m + v_m)^T N_m}{n_m} \to (u_m + v_m)^T s
$$
in probability.

Similarly, we have

$$
\frac{(u_p + v_p)^T N_p}{n_p} \to (u_p + v_p)^T s
$$
in probability.

So

$$
w^2 (u_m + v_m)^T N_m + (1 - w)^2 (u_p + v_p)^T N_p
$$

$$
- [n_m w^2 (u_m + v_m)^T s + n_p (1 - w)^2 (u_p + v_p)^T s]
$$

$$
\to 0 \text{ in probability,}
$$

$$
w(u_m - v_m)^T N_m \pm (1 - w)(u_p - v_p)^T N_p
$$

$$
- [n_m w(u_m - v_m)^T s \pm n_p (1 - w)(u_p - v_p)^T s]
$$

$$
\to 0 \text{ in probability,}
$$

$$
\frac{n_m w^2 [(u_m + v_m)^T s - ((u_m - v_m)^T s)^2]}{\sqrt{n_m w^2 (u_m + v_m)^T s + n_p (1 - w)^2 (u_p + v_p)^T s}}
$$

$$
\to N(0, 1) \text{ in law.}
$$

Hence for arbitrary $w \in [0,1]$, we have

$$
\frac{w(u_m - v_m)^T N_m \pm (1 - w)(u_p - v_p)^T N_p}{\sqrt{w^2 (u_m + v_m)^T N_m + (1 - w)^2 (u_p + v_p)^T N_p}}
$$

$$
- \frac{n_m w(u_m - v_m)^T s \pm n_p (1 - w)(u_p - v_p)^T s}{\sqrt{n_m w^2 (u_m + v_m)^T s + n_p (1 - w)^2 (u_p + v_p)^T s}}
$$

$$
\to N \left( 0, \frac{n_m w^2 [(u_m + v_m)^T s - ((u_m - v_m)^T s)^2]}{n_m w^2 (u_m + v_m)^T s + n_p (1 - w)^2 (u_p + v_p)^T s} \right).
$$

In the remainder of this section, the population is assumed to be in Hardy-Weinberg equilibrium. From the detailed expressions of $s_1, \ldots, s_{15}$ in Zhou et al. [2006], we have

$$
(u_m + v_m)^T s = 2 g_8' + 2 \Delta(1 - 2 g)(1 + \theta_f - \theta_m)
$$

$$
+ 2 \delta R \delta^2 (0_m - 1) + \frac{I(1 - 2 g)(1 - 2 \theta)}{\phi},
$$

$$
(u_m - v_m)^T s = \delta \left[ \Delta(1 - 2 \theta) + \frac{I(1 + \theta_f - \theta_m)}{\phi} \right],
$$

$$
(u_p + v_p)^T s = 2 g_8' + 2 \Delta(1 - 2 g)(1 + \theta_m - \theta_f)
$$

$$
+ 2 \delta R \delta^2 (\theta_f - 1) - \delta I(1 - 2 g)(1 - 2 \theta),
$$

$$
(u_p - v_p)^T s = \delta \left[ \Delta(1 - 2 \theta) + \frac{I(\theta_f - 1 - \theta_m)}{\phi} \right],
$$

where $R = \phi_{D/D} - \phi_{D/d} - \phi_{d/D} + \phi_{d/d}$ is the difference between two homozygote risks and two heterozygote risks, $\Delta = (p(2\phi_{D/D} - \phi_{D/d} - \phi_{d/D}) + q(\phi_{D/d} + \phi_{d/D} - 2\phi_{d/d}))/2\phi$. In fact, $\Delta$ is the difference between two ratios $P(D\text{~affected child})/P(D)$ and $P(d\text{~affected child})/P(d)$, where $P(D\text{~affected child})$ represents the probability that a chromosome of an affected child has a disease allele $D$ at a DSL, and the other probability $P(d\text{~affected child})$ is similarly defined. The ratio difference $\Delta$ is a positive quantity according to the relative magnitude of the four risk parameters.

Particularly, when $l = 0$ and $\theta_f = \theta_m$, we have from equation (8)

$$
T_w = \frac{n_m w + n_p (1 - w)}{\sqrt{(n_m + n_p)(n_m w^2 + n_p (1 - w)^2)} \mu} \to N(0, \sigma^2),
$$

(9)
where
\[
\mu = \sqrt{n_m + n_p} \frac{\delta \Delta \sqrt{\phi(1 - 2\theta)}}{\sqrt{2g' \phi + \delta \phi \Delta(1 - 2g) + 2\theta \delta^2(\theta - 1)}},
\]
\[
\sigma^2 = 1 - \frac{\phi \delta^2 \Delta^2 (1 - 2\theta)^2}{2g' \phi + \delta \phi \Delta(1 - 2g) + 2\theta \delta^2(\theta - 1)}.
\]
When \( \theta_j = \theta_m = 0.5 \) and \( I = 0 \), we have \( T_w \rightarrow N(0, 1) \). It implies that \( T_w \) with any \( w \in [0, 1] \) can be used to test for linkage in the case of \( I = 0 \). When \( \theta_j = \theta_m, I = 0 \), and \( w = 0.5 \), we have 1-TDT

\[
\text{UNBIASED ESTIMATOR OF THE VARIANCE}
\]

\[
\mathbb{E}(1) = \frac{\sqrt{2} \Phi(n_m - n_p)}{\sqrt{(n_m + n_p)(4g' \phi^2 + 2\delta \phi \Delta^2 + R \phi \delta^2)}}.
\]

So the mean of 1-TDT under the null hypothesis of no linkage could be biased from zero, unless \( n_m = n_p \) or \( \delta = 0 \).

**APPENDIX C**

**UNBIASED ESTIMATOR OF THE VARIANCE**

First, we have
\[
\mathbb{E}(w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C})) = \frac{n_m n_p}{n_m + n_p} \mathbb{E}(I_{M<C} - I_{M>C}) + (I_{F<C} - I_{F>C})
\]
and
\[
\mathbb{E}(w_0(N_{M<C} - N_{M>C}) - (1 - w_0)(N_{F<C} - N_{F>C})) = \frac{n_m n_p}{n_m + n_p} \mathbb{E}(I_{M<C} - I_{M>C}) - (I_{F<C} - I_{F>C})
\]
for a general population that does not require the assumption of Hardy-Weinberg equilibrium. Furthermore, we have
\[
\begin{align*}
\mathbb{E}(I_{M<C} &= P(M = 1, C = 2|\text{child is affected}) \\
+ P(M = 0, C = 1|\text{child is affected}) \\
= P(F = 2, M = 1, C = 2|\text{child is affected}) \\
+ P(F = 1, M = 1, C = 2|\text{child is affected}) \\
+ P(F = 2, M = 0, C = 1|\text{child is affected}) \\
+ P(F = 1, M = 0, C = 1|\text{child is affected}) \\
= s_1 + s_5 + s_8 + s_{13}.
\end{align*}
\]

Similarly, we have \( \mathbb{E}(I_{M>C} = s_4 + s_7 + s_{11} + s_{14}, \mathbb{E}(I_{F<C} = s_2 + s_5 + s_9 + s_{14}), \mathbb{E}(I_{F>C} = s_3 + s_7 + s_{10} + s_{13}) \). So we have
\[
\begin{align*}
\mathbb{E}[(I_{M<C} - I_{M>C}) + (I_{F<C} - I_{F>C})] &= s_1 - s_3 + s_2 \\
- s_4 + 2(s_5 - s_7) + s_8 - s_{10} - s_9 - s_{11}, \\
\mathbb{E}[(I_{M<C} - I_{M>C}) - (I_{F<C} - I_{F>C})] &= s_1 - s_2 + s_3 \\
- s_4 + s_8 - s_9 - s_{10} + s_11 + 2(s_{13} - s_{14}).
\end{align*}
\]

Under the null hypothesis of no linkage, we have
\( s_1 = s_3, s_2 = s_4, s_5 = s_7, s_8 = s_{10}, s_9 = s_{11} \). When the Hardy–Weinberg law holds among the parents in the source population, we have \( s_1 = s_3, s_2 = s_4, s_5 = s_7, s_8 = s_{10}, s_9 = s_{11} \) under the null hypothesis of no association. So under the null hypothesis of no linkage/association, we have
\[
\begin{align*}
\mathbb{E}(w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C})) &= 0. \\
\text{Under the null hypothesis of no LD, } [w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C})]^2 \text{ is shown to be an unbiased estimator of the variance of } w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C}). \text{ Let } A = \mathbb{E}(I_{M<C} - I_{M>C}) \text{[no LD]} - \mathbb{E}(I_{F<C} - I_{F>C}) \text{[no LD]}, \text{ then}
\mathbb{E}(w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C}))^2
\end{align*}
\]
\[
= \mathbb{E}(w_0^2 N_{M\neq C} + (1 - w_0)^2 N_{F\neq C}) \\
- 2w_0(1 - w_0)n_m n_p A^2 \\
= \mathbb{E}(w_0^2 N_{M\neq C} + (1 - w_0)^2 N_{F\neq C}) - n_m n_p (n_m + n_p)^{-1} A^2 \\
= \mathbb{E}(w_0^2 N_{M\neq C} + (1 - w_0)^2 N_{F\neq C}) + (n_m + n_p)^{-1} \\
(N_{M<C} - N_{M>C})(N_{F<C} - N_{F>C}).
\]

So \( w_0^2 N_{M\neq C} + (1 - w_0)^2 N_{F\neq C} + (n_m + n_p)^{-1}(N_{M<C} - N_{M>C})(N_{F<C} - N_{F>C}) \) is an unbiased estimator of the variance of \( w_0(N_{M<C} - N_{M>C}) + (1 - w_0)(N_{F<C} - N_{F>C}) \) under the null hypothesis of no
linkage where Hardy–Weinberg equilibrium needs not to be assumed, or under the null hypothesis of no association where the Hardy–Weinberg law is taken among the parents in the source population.

When \( \theta_f = \theta_m \) and the population mating is symmetry, we have \( s_1 = s_2, s_3 = s_4, s_8 = s_9, s_{10} = s_{11}, s_{13} = s_{14} \) under the null hypothesis of no imprinting. So under the null hypothesis, we have

\[
E[w_0(N_{M< C} - N_{M> C}) - (1 - w_0)(N_{F< C} - N_{F> C})] = 0.
\]

Similarly, we can verify that \( w_0^2 N_{M\neq C} + (1 - w_0)^2 N_{F\neq C} - (n_m + n_p)^{-1}(N_{M< C} - N_{M> C})(N_{F< C} - N_{F> C}) \) is an unbiased estimator of the variance of \( w_0(N_{M< C} - N_{M> C}) - (1 - w_0)(N_{F< C} - N_{F> C}) \) under the null hypothesis of no imprinting. Note that Hardy–Weinberg equilibrium needs not to be assumed here.

Employing the same principle, we can also verify that \( \text{Var}_0(T) \) in equation (4) is an unbiased estimator of the variance of \( T \) under the null hypothesis of no LD/imprinting.