*Am. J. Hum. Genet. 64:1177–1185, 1999*

# A Note on Power Approximations for the Transmission/Disequilibrium Test

Michael Knapp

Institute for Medical Statistics, University of Bonn, Bonn

## Summary

**The transmission/disequilibrium test (TDT) is a popular method for detection of the genetic basis of a disease. Investigators planning such studies require computation of sample size and power, allowing for a general genetic model. Here, a rigorous method is presented for obtaining the power approximations of the TDT for samples consisting of families with either a single affected child or affected sib pairs. Power calculations based on simulation show that these approximations are quite precise. By this method, it is also shown that a previously published power approximation of the TDT is erroneous.**

## Introduction

Whereas linkage analysis has been successfully used to localize genes that cause monogenic diseases, this method was less successful in the detection of genetic factors for complex diseases such as asthma and psychiatric disorders. Therefore, it has been argued (Lander 1996; Risch and Merikangas 1996) that the future of the genetics of complex diseases will require large-scale testing by association analysis. Risch and Merikangas (1996) compared the mean test of affected-sib-pair (ASP) linkage analysis (Blackwelder and Elston 1985) with the transmission/disequilibrium test (TDT; Spielman et al. 1993), with respect to the power of these methods to detect genes of modest effect. In their work, they assumed a multiplicative relationship for the genotypic relative risk and calculated sample sizes required to obtain a prescribed power under this mode of inheritance (MOI). Recently, Camp (1997) tried to extend these sample-size calculations to general MOIs. Her approach for approximating the power of the TDT, however, is technically inadequate, and, therefore, the sample sizes obtained by Camp (1997) are misleading. The purpose of the present paper is to derive two valid approximations for the power of the TDT for general MOIs. It will be shown that the second of these approximations coincides with the approach of Risch and Merikangas (1996) in case of a multiplicative MOI and nuclear families with a single affected child. Power calculations based on simulation reveal that the first approximation proposed in the present paper is even more precise. Therefore, this approximation can be useful not only to compare the TDT with ASP linkage analysis but also to illustrate the potential resolution of an intended association study, as will be shown by a real-data example.

Throughout this paper, the "classic" TDT situation is considered, which is characterized by (i) a biallelic marker locus, (ii) a qualitative trait (disease), and (iii) the availability of both parents. A recent overview by Schaid (1998) has discussed several extensions of the TDT.

## Methods

### *A General Approach for Power Approximations for the TDT*

This section presents a rigorous derivation of two different general approximations for the power of the TDT, which will then be applied, in the next section, both to nuclear families with a single affected child and to families with two affected children. The TDT can be described in the following way: The marker genotypes of the parents together with the marker genotype(s) of the child (children) constitute the *type* of the family. Clearly, the number of possible types of families depends on the ascertainment scheme. Assume that there are $(k + 1)$ different types of families. For $1 \leq j \leq k + 1$, let $s_j$ denote the probability that a family is of type $j$. These probabilities depend on the genetic model. For example, in the case of nuclear families each with a single affected child, $s_j$ may be calculated for a specified allele frequency $p$ of the putative disease allele $A$, relative risks $\Psi_1$ and $\Psi_2$ (with $\Psi_i$ denoting the relative risk of an individual carrying $i$ $A$ alleles, compared with that of an individual carrying none), and the additional assumption of Hardy-

Received November 4, 1998; accepted for publication January 21, 1999; electronically published March 5, 1999.

Address for correspondence and reprints: Dr. Michael Knapp, Institute for Medical Statistics, University of Bonn, Sigmund-Freud-Strasse 25, D-53105 Bonn, Germany. E-mail: knapp@imsdd.meb.uni-bonn.de


0002-9297/99/6404-0030$02.00

Weinberg equilibrium in the parental generation. Let $Z_{jn}$ be the random variable that denotes the number of families of type $j$ in a sample of size $n$. Then, $\boldsymbol{Z}_n := (Z_{1n},...,Z_{(k+1)n})$ is a $(k+1)$-dimensional multinomially distributed random variable with parameters $(n,s_1,...,s_{k+1})$, where $0 \leq s_j \leq 1$ for $1 \leq j \leq (k+1)$ and $s_{k+1} = 1 - \Sigma_{j=1}^{k} s_j$. For family type $j$, let $u_j$ denote the number of alleles $A$ transmitted from heterozygous parents to their affected offspring and let $v_j$ denote the number of alleles $A$ not transmitted from heterozygous parents to their affected children. For example, if family type $j$ is an $AB \times AA$ parental mating with one affected child being $AA$ and a second affected child being $AB$, then $u_j = 1$ and $v_j = 1$. Finally, assume that family type $(k+1)$ combines all family types that are uninformative (i.e., both parents are homozygous) for the TDT and define

$$X_n := \frac{\sum_{j=1}^{k} u_j \cdot Z_{jn} - \sum_{j=1}^{k} v_j \cdot Z_{jn}}{\sqrt{\sum_{j=1}^{k} u_j \cdot Z_{jn} + \sum_{j=1}^{k} v_j \cdot Z_{jn}}} .$$

Then, the test statistic of the TDT is $X_n^2$.

It will now be shown that the distribution of $X_n$ can be approximated by a normal distribution. For this purpose, let $\boldsymbol{T}_n := (T_{1n},...,T_{kn})$ with $T_{jn} := Z_{jn}/n$ for $1 \leq j \leq k$ denote the maximum-likelihood estimator of $\boldsymbol{s} := (s_1,...,s_k)$. It is well known (see Johnson and Kotz 1972, p. 38) that the asymptotic distribution of $\sqrt{n} \cdot (\boldsymbol{T}_n - \boldsymbol{s})$ is $k$-variate normal with mean 0 and dispersion matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq k}$, where $\sigma_{ij} = s_i \cdot (1 - s_i)$ for $i = j$ and $\sigma_{ij} = -s_i \cdot s_j$ for $i \neq j$. Now let $\boldsymbol{u} := (u_j)_{1 \leq j \leq k}$ and let $\boldsymbol{v} := (v_j)_{1 \leq j \leq k}$ and define $U_n := \boldsymbol{u}^T \cdot \boldsymbol{T}_n$ and $V_n := \boldsymbol{v}^T \cdot \boldsymbol{T}_n$. Then the asymptotic distribution of $\sqrt{n} \cdot [(U_n,V_n) - (\boldsymbol{u}^T \cdot \boldsymbol{s}, \boldsymbol{v}^T \cdot \boldsymbol{s})]$ is bivariate normal with mean 0 and dispersion matrix $\Sigma^* = (\sigma_{ij}^*)_{1 \leq i,j \leq 2}$, where

$$\sigma_{11}^* = \boldsymbol{u}^T \cdot \sum \cdot \boldsymbol{u} = \sum_{i=1}^{k} u_i \cdot \left( s_i \cdot u_i - \sum_{j=1}^{k} s_i \cdot s_j \cdot u_j \right) = \sum_{i=1}^{k} u_i^2 \cdot s_i - \left( \sum_{i=1}^{k} u_i \cdot s_i \right)^2 ,$$

$$\sigma_{12}^* = \boldsymbol{u}^T \cdot \sum \cdot \boldsymbol{v} = \sum_{i=1}^{k} u_i \cdot v_i \cdot s_i - \left( \sum_{i=1}^{k} u_i \cdot s_i \right) \times \left( \sum_{i=1}^{k} v_i \cdot s_i \right) ,$$

$$\sigma_{22}^* = \boldsymbol{v}^T \cdot \sum \cdot \boldsymbol{v} = \sum_{i=1}^{k} v_i^2 \cdot s_i - \left( \sum_{i=1}^{k} v_i \cdot s_i \right)^2 .$$

*First approximation.*—Let $g(x_1,x_2) := (x_1 - x_2)/\sqrt{x_1 + x_2}$. Then,

$$g(U_n,V_n) = \frac{U_n - V_n}{\sqrt{U_n + V_n}} = \frac{1}{\sqrt{n}} \cdot X_n ,$$

and it follows, from Rao (1973, p. 387), that the asymptotic distribution of

$$\sqrt{n} \cdot \left( \frac{U_n - V_n}{\sqrt{U_n + V_n}} - \frac{\boldsymbol{u}^T \cdot \boldsymbol{s} - \boldsymbol{v}^T \cdot \boldsymbol{s}}{\sqrt{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}} \right)$$

is normal with mean 0 and variance

$$\sigma_{A1}^2 = \sum_{1 \leq i,j \leq 2} \sigma_{ij}^* \cdot \frac{\partial g}{\partial x}_{j \mid (x_1,x_2) = (\boldsymbol{u}^T \cdot \boldsymbol{s}, \boldsymbol{v}^T \cdot \boldsymbol{s})} \times \frac{\partial g}{\partial x}_{j \mid (x_1,x_2) = (\boldsymbol{u}^T \cdot \boldsymbol{s}, \boldsymbol{v}^T \cdot \boldsymbol{s})} .$$

Thus, the distribution of $X_n$ can be approximated by a normal distribution with mean

$$\mu_{X_n} := \sqrt{n} \cdot \frac{\boldsymbol{u}^T \cdot \boldsymbol{s} - \boldsymbol{v}^T \cdot \boldsymbol{s}}{\sqrt{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}} \tag{1}$$

and variance $\sigma_{A1}^2$. With the abbreviations

$$e_1 := \boldsymbol{u}^T \cdot \boldsymbol{s} , \tag{2}$$

$$e_2 := \boldsymbol{v}^T \cdot \boldsymbol{s} , \tag{3}$$

$$d_1 := \sum_{i=1}^{k} (u_i - v_i)^2 \cdot s_i , \tag{4}$$

$$d_2 := \sum_{i=1}^{k} (u_i + v_i)^2 \cdot s_i , \tag{5}$$

$$d_{1,2} := \sum_{i=1}^{k} (v_i - u_i) \cdot (v_i + u_i) \cdot s_i , \tag{6}$$

the Appendix shows that

$$\sigma_{A1}^2 = \frac{d_1 - (e_1 - e_2)^2/4}{(e_1 + e_2)} + \frac{(e_1 - e_2) \cdot d_{1,2}}{(e_1 + e_2)^2} + \frac{1}{4} \cdot \frac{(e_1 - e_2)^2 \cdot d_2}{(e_1 + e_2)^3} . \tag{7}$$

Now let $\chi^2_{1,x}$ denote the $x$-quantile of a $\chi^2$ distribution with 1 df and let $z_x$ denote the $x$-quantile of a standard normal distribution. Then,

$$\begin{aligned}\text{Power of the TDT} &= P(X_n^2 > \chi^2_{1,1-\alpha}) \\ &= P(X_n < -z_{1-\alpha/2}) + P(X_n > z_{1-\alpha/2}) \\ &\approx P\left(Z < \frac{-z_{1-\alpha/2} - \mu_{X_n}}{\sigma_{A1}}\right) \\ &\quad + P\left(Z > \frac{z_{1-\alpha/2} - \mu_{X_n}}{\sigma_{A1}}\right) ,\end{aligned} \tag{8}$$

with $Z$ being a standard normal–distributed random variable.

The TDT is usually performed as a two-sided test of the null hypothesis of no linkage—that is, $H_0$ is rejected if there is excess transmission of either allele $A$ or allele $B$. Risch and Merikangas (1996), however, consider a one-sided version of the TDT—that is, $H_0$ is rejected only if allele $A$ has been transmitted more often than allele $B$. The power of this one-sided TDT is approximated as

$$\text{Power of one-sided TDT} \approx P\left(Z > \frac{z_{1-\alpha} - \mu_{X_n}}{\sigma_{A1}}\right) . \tag{9}$$

If allele $A$ is positively associated with the disease and $\alpha$ is small, then the power of the one-sided TDT with a type I error probability of $\alpha$ is very near the power of the two-sided TDT with a type I error of $2\alpha$.

*Second approximation.*—Now let

$$\begin{aligned}h(x_1,x_2) &:= (x_1 - x_2)/\sqrt{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}} \\ &= (x_1 - x_2)/\sqrt{e_1 + e_2} .\end{aligned}$$

Then,

$$\begin{aligned}h(U_n,V_n) &= \frac{U_n - V_n}{\sqrt{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}} \\ &= \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{U_n + V_n}{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}} \cdot X_n .\end{aligned}$$

It follows that the asymptotic distribution of

$$\sqrt{n} \cdot \left(\frac{U_n - V_n}{\sqrt{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}} - \frac{\boldsymbol{u}^T \cdot \boldsymbol{s} - \boldsymbol{v}^T \cdot \boldsymbol{s}}{\sqrt{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}}\right)$$

is normal, with mean 0 and variance

$$\sigma^2_{A2} = \frac{\sigma^*_{11} - 2\sigma^*_{12} + \sigma^*_{22}}{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}} = \frac{d_1 - (e_1 - e_2)^2}{e_1 + e_2} . \tag{10}$$

Thus, the distribution of

$$\sqrt{\frac{U_n + V_n}{\boldsymbol{u}^T \cdot \boldsymbol{s} + \boldsymbol{v}^T \cdot \boldsymbol{s}}} \cdot X_n$$

can be approximated by a normal distribution with mean $\mu_{X_n}$ and variance $\sigma^2_{A2}$. Since $\lim_{n\to\infty}\text{Var}(U_n + V_n) = 0$, this normal approximation may be used for $X_n$ itself. A second approximation of the power of the TDT is then obtained by replacing $\sigma_{A1}$ by $\sigma_{A2}$ in approximations (8) and (9); that is,

$$\begin{aligned}\text{Power of the TDT} &\approx P\left(Z < \frac{-z_{1-\alpha/2} - \mu_{X_n}}{\sigma_{A2}}\right) \\ &\quad + P\left(Z > \frac{z_{1-\alpha/2} - \mu_{X_n}}{\sigma_{A2}}\right) ,\end{aligned} \tag{11}$$

and

$$\text{Power of one-sided TDT} \approx P\left(Z > \frac{z_{1-\alpha} - \mu_{X_n}}{\sigma_{A2}}\right) . \tag{12}$$

### *Singletons*

In the case of nuclear families each with a single affected child, there are seven informative family types, which are listed in table 1, together with their corresponding probabilities (i.e., $s_j$) under the assumption of Hardy-Weinberg equilibrium in the parental generation.

**Table 1**

**Classification of Family Types for Nuclear Families Each with a Single Affected Child**

| | Family Type[a] | | | | |
|---|---|---|---|---|---|
| $j$ | Parental Mating | Genotype of Affected Child | $s_j$ | $u_j$ | $v_j$ |
| 1 | $AA \times AB$ | $AA$ | $2p^3q\Psi_2/R$ | 1 | 0 |
| 2 | $AA \times AB$ | $AB$ | $2p^3q\Psi_1/R$ | 0 | 1 |
| 3 | $AB \times AB$ | $AA$ | $p^2q^2\Psi_2/R$ | 2 | 0 |
| 4 | $AB \times AB$ | $AB$ | $2p^2q^2\Psi_1/R$ | 1 | 1 |
| 5 | $AB \times AB$ | $BB$ | $p^2q^2/R$ | 0 | 2 |
| 6 | $AB \times BB$ | $AB$ | $2pq^3\Psi_1/R$ | 1 | 0 |
| 7 | $AB \times BB$ | $BB$ | $2pq^3/R$ | 0 | 1 |

[a] $R := \Psi_2 p^2 + \Psi_1 2pq + q^2$.

When $s_j$,$u_j$ and $v_j$ of table 1 are inserted into formulas (2)–(6), it follows that

$$e_1 + e_2 = \frac{2pq}{R}(p\Psi_2 + \Psi_1 + q) \ , \qquad (13)$$

$$e_1 - e_2 = \frac{2pq}{R}[p\Psi_2 + (1 - 2p)\Psi_1 - q] \ , \qquad (14)$$

$$d_1 = \frac{2pq}{R}[(1 - q^2)\Psi_2 + (1 - 2pq)\Psi_1 + (1 - p^2)] \ , \qquad (15)$$

$$d_2 = \frac{2pq}{R}[(1 - q^2)\Psi_2 + (1 + 2pq)\Psi_1 + (1 - p^2)] \ , \qquad (16)$$

$$d_{1,2} = \frac{2pq}{R}[-(1 - q^2)\Psi_2 + (p - q)\Psi_1 + (1 - p^2)] \ . \qquad (17)$$

Now, formulas (13)–(17) can be used to calculate $\mu_{X_n}$, $\sigma^2_{A1}$, and $\sigma^2_{A2}$ according to formulas (1), (7), and (10) and, finally, to obtain power approximations for the TDT, according to formulas (8), (9), (11), and (12). Note that

$$d_1 - (e_1 - e_2)^2 = \frac{2pq}{R^2}(R^2 + 2pq\Psi_2 + q^2\Psi_1 + p^2\Psi_2\Psi_1) \ . \qquad (18)$$

With formulas (13), (14), and (18), it can be seen that the second power approximation, formula (11), for the TDT with singletons is equivalent to the approach already described by Baur and Knapp (1997, p. 169).

Risch and Merikangas (1996) considered the special case of a multiplicative model—that is, $\Psi_1 = \gamma$ and $\Psi_2 = \Psi_1^2 = \gamma^2$. For this kind of disease model, formulas (13), (14), and (18) further simplify to

$$e_1 + e_2 = 2h,$$

$$e_1 - e_2 = 2h \cdot \frac{\gamma - 1}{\gamma + 1} \ ,$$

$$d_1 - (e_1 - e_2)^2 = 2h \cdot \left[1 - h \cdot \frac{(\gamma - 1)^2}{(\gamma + 1)^2}\right],$$

with $h := pq \cdot (\gamma + 1)/(p\gamma + q)$ denoting the probability that a parent in a family with a single affected child is heterozygous for allele *A*. Therefore,

$$\mu_{X_n} = \sqrt{2n} \cdot \sqrt{h} \cdot \frac{\gamma - 1}{\gamma + 1}, \ \sigma^2_{A2} = 1 - h \cdot \frac{(\gamma - 1)^2}{(\gamma + 1)^2} \ . \qquad (19)$$

Comparison of formula (19) with note 6 in Risch and Merikangas's (1996) work reveals that their power approximation for singletons is identical to the second approximation, formula (12), for the one-sided TDT in the case of a multiplicative model.

### *Sib Pairs*

The informative family types and corresponding probabilities for families each with an ASP are given in table 2. Again, Hardy-Weinberg equilibrium in the parental generation was assumed for derivation of $s_j$. Additionally, it was assumed that disease occurrences in sib pairs are independent, conditional on their genotypes at the trait locus in question. When $s_j$,$u_j$ and $v_j$ from table 2 are inserted into formulas (2)–(6), elementary algebra shows that

$$e_1 + e_2 = \frac{pq}{R}[2p^2(\Psi_2 + \Psi_1)^2 + pq(\Psi_2 + 2\Psi_1 + 1)^2 + 2q^2(\Psi_1 + 1)^2] \ , \qquad (20)$$

$$e_1 - e_2 = \frac{pq}{R}[2p^2(\Psi_2^2 - \Psi_1^2) + pq(\Psi_2^2 + 2\Psi_2\Psi_1 - 2\Psi_1 - 1) + 2q^2(\Psi_1^2 - 1)] \ , \qquad (21)$$

**Table 2**

**Classification of Family Types for Nuclear Families Each with an ASP**

| | Family Type[a] | | | | |
|---|---|---|---|---|---|
| $j$ | Parental Mating | Genotypes of Affected Children | $s_j$ | $u_j$ | $v_j$ |
| 1 | $AA \times AB$ | *AA*, *AA* | $p^3q\Psi^2_2/R$ | 2 | 0 |
| 2 | $AA \times AB$ | *AA*, *AB* | $2p^3q\Psi_2\Psi_1/R$ | 1 | 1 |
| 3 | $AA \times AB$ | *AB*, *AB* | $p^3q\Psi^2_1/R$ | 0 | 2 |
| 4 | $AB \times AB$ | *AA*, *AA* | $\frac{1}{4}p^2q^2\Psi^2_2/R$ | 4 | 0 |
| 5 | $AB \times AB$ | *AA*, *AB* | $p^2q^2\Psi_2\Psi_1/R$ | 3 | 1 |
| 6 | $AB \times AB$ | *AA*, *BB* | $\frac{1}{2}p^2q^2\Psi_2/R$ | 2 | 2 |
| 7 | $AB \times AB$ | *AB*, *AB* | $p^2q^2\Psi^2_1/R$ | 2 | 2 |
| 8 | $AB \times AB$ | *AB*, *BB* | $p^2q^2\Psi_1/R$ | 1 | 3 |
| 9 | $AB \times AB$ | *BB*, *BB* | $\frac{1}{4}p^2q^2/R$ | 0 | 4 |
| 10 | $AB \times BB$ | *AB*, *AB* | $pq^3\Psi^2_1/R$ | 2 | 0 |
| 11 | $AB \times BB$ | *AB*, *BB* | $2pq^3\Psi_1/R$ | 1 | 1 |
| 12 | $AB \times BB$ | *BB*, *BB* | $pq^3/R$ | 0 | 2 |

[a] $R := \Psi^2_2p^4 + (\Psi_2 + \Psi_1)^2p^3q + \Psi^2_1 2p^2q^2 + (\Psi_2 + 2\Psi_1 + 1)^2\frac{1}{4}p^2q^2 + (\Psi_1 + 1)^2pq^3 + q^4$.

$$d_1 = \frac{4pq}{R}[p^2(\Psi_2^2 + \Psi_1^2) + pq(\Psi_2^2 + \Psi_2\Psi_1 + \Psi_1 + 1) + q^2(\Psi_1^2 + 1)] , \quad (22)$$

$$d_2 = \frac{4pq}{R}[p^2(\Psi_2 + \Psi_1)^2 + pq(\Psi_2 + 2\Psi_1 + 1)^2 + q^2(\Psi_1 + 1)^2] ,$$

$$d_{1,2} = -\frac{4pq}{R}[p^2(\Psi_2^2 - \Psi_1^2) + pq(\Psi_2^2 + 2\Psi_2\Psi_1 - 2\Psi_1 - 1) + q^2(\Psi_1^2 - 1)] .$$

For the special case of the multiplicative model considered by Risch and Merikangas, let

$$h := \frac{pq \cdot (\gamma + 1)^2}{2 \cdot (p\gamma + q)^2 + pq \cdot (\gamma - 1)^2}$$

denote the probability that any given parent in a family with an ASP is heterozygous for allele $A$. Some tedious but elementary algebra shows that, for this kind of disease model, formulas (20)–(22) simplify to

$$e_1 + e_2 = 4h ,$$

$$e_1 - e_2 = 4h\frac{\gamma - 1}{\gamma + 1} ,$$

$$d_1 = 4h \cdot \left[1 + 3h \cdot \frac{(\gamma - 1)^2}{(\gamma + 1)^2}\right] + \frac{2pq \cdot (\gamma - 1)^2}{R} \cdot (p^2 \cdot \gamma^2 + q^2) ,$$

and, therefore,

$$d_1 - (e_1 - e_2)^2 = 4h \cdot \left[1 - h \cdot \frac{(\gamma - 1)^2}{(\gamma + 1)^2}\right] + \frac{2pq \cdot (\gamma - 1)^2}{R} \cdot (p^2 \cdot \gamma^2 + q^2) . \quad (23)$$

Now, comparison of formula (23) with note 6 in Risch and Merikangas's (1996) work shows that their power approximation for sib pairs is not identical to the second approximation, formula (12), for the one-sided TDT in the case of a multiplicative model. Since the second term on the right side of formula (23) is always >0 for $\gamma > 1$, the variance used by Risch and Merikangas (1996) is smaller than $\sigma_{A2}^2$, so that their sample sizes necessary to gain a prescribed power are smaller than the sample sizes calculated from the second approximation, formula (12).

### *Simulation Study*

The precision of the proposed power approximation was checked by simulations. For these simulations, the same modes of inheritance (MOI) as were used by Camp (1997) were considered: (1) the multiplicative model (i.e., $\Psi_1 = \gamma$ and $\Psi_2 = \gamma^2$), (2) the additive model (i.e., $\Psi_1 = \gamma$ and $\Psi_2 = 2\gamma$), (3) the recessive model (i.e., $\Psi_1 = 1$ and $\Psi_2 = \gamma$), and (4) the dominant model (i.e., $\Psi_1 = \Psi_2 = \gamma$). With $f_i$ denoting the penetrances, it could be argued that an additive MOI requires $f_1 = (f_0 + f_2)/2$, which would imply that $\Psi_2 = 2 \cdot \Psi_1 - 1$, but, for the sake of comparability, the present article paper adopts Camp's (1997) interpretation of additive MOI. For each combination of MOI, risk parameter $\gamma$ (with $\gamma \in \{1.5,2,4\}$), and population frequency $p$ of allele $A$ (with $p \in \{.01,.1,.5,.8\}$), the sample sizes $n$ necessary to gain 80% power for the two-sided TDT with $\alpha = 10^{-7}$ were calculated according to the power approximations (8) and (11). Next, $10^5$ replicates of these sample sizes were generated. For each replicate, the TDT statistic was calculated, and the true power was estimated as the proportion of replicates being significant at $\alpha = 10^{-7}$. With $10^5$ replicates, the standard error of this power estimate is ≈.0013. Therefore, the first two digits of this estimate can be expected to be correct and should equal .80 in the case that power approximations (8) and (11) are very precise.

## Results

### *Singletons*

Table 3 contains the results of the simulation for families each with a single affected child. The simulation reveals that the second approximation, formula (11), tends to underestimate the power of the TDT and, therefore, that it overestimates the sample size necessary to gain 80% power. This effect is most pronounced for $\gamma = 4$. On the other hand, the simulated power for sample sizes obtained by the first approximation, formula (8), matches very well with the expected value of .80. Comparison of the sample sizes given in table 3 with those given by Camp (1997, table 3) shows that the sample sizes calculated by Camp (1997) are much too

## Table 3

**Sample Size Necessary to Gain 80% Power in TDT with Singletons ($\alpha = 10^{-7}$), According to First Approximation (Formula [8]) and Second Approximation (Formula [11])**

| | Sample Size Necessary to Gain 80% Power (True Power of TDT[a]) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Multiplicative MOI | | Additive MOI | | Recessive MOI | | Dominant MOI | |
| $\gamma_{AND}$ $p$ | First Approximation | Second Approximation | First Approximation | Second Approximation | First Approximation | Second Approximation | First Approximation | Second Approximation |
| 4.0: | | | | | | | | |
| .01 | 1,056 (.79) | 1,097 (.83) | 1,095 (.80) | 1,136 (.83) | 4,344,070 (.80) | 4,344,285 (.80) | 1,115 (.80) | 1,156 (.83) |
| .10 | 146 (.80) | 150 (.82) | 194 (.80) | 197 (.82) | 5,631 (.80) | 5,647 (.80) | 231 (.80) | 235 (.82) |
| .50 | 101 (.80) | 103 (.82) | 218 (.80) | 220 (.81) | 207 (.80) | 209 (.81) | 696 (.80) | 698 (.80) |
| .80 | 216 (.80) | 222 (.83) | 553 (.80) | 559 (.81) | 259 (.80) | 264 (.82) | 9,384 (.80) | 9,393 (.80) |
| 2.0: | | | | | | | | |
| .01 | 5,755 (.80) | 5,820 (.81) | 5,755 (.80) | 5,820 (.81) | 38,654,522 (.80) | 38,654,716 (.80) | 5,947 (.80) | 6,012 (.81) |
| .10 | 689 (.80) | 695 (.81) | 689 (.80) | 695 (.81) | 45,071 (.80) | 45,089 (.80) | 949 (.80) | 954 (.80) |
| .50 | 338 (.80) | 340 (.81) | 338 (.80) | 340 (.81) | 957 (.80) | 959 (.80) | 1,839 (.80) | 1,841 (.80) |
| .80 | 634 (.80) | 640 (.81) | 634 (.80) | 640 (.81) | 851 (.80) | 855 (.81) | 21,998 (.80) | 22,006 (.80) |
| 1.5: | | | | | | | | |
| .01 | 19,233 (.80) | 19,310 (.80) | 18,733 (.80) | 18,811 (.80) | 154,174,890 (.80) | 154,174,896 (.80) | 19,755 (.80) | 19,831 (.80) |
| .10 | 2,210 (.80) | 2,217 (.80) | 1,755 (.80) | 1,763 (.80) | 174,694 (.80) | 174,713 (.80) | 2,897 (.80) | 2,903 (.80) |
| .50 | 947 (.80) | 949 (.80) | 464 (.80) | 466 (.80) | 3,099 (.80) | 3,100 (.80) | 4,568 (.80) | 4,570 (.80) |
| .80 | 1,658 (.80) | 1,663 (.80) | 698 (.80) | 703 (.81) | 2,356 (.80) | 2,360 (.80) | 50,826 (.80) | 50,834 (.80) |

[a] Estimated by simulation with $10^5$ replicates.

## Table 4

**Sample Size Necessary to Gain 80% Power in TDT with Sib Pairs ($\alpha = 10^{-7}$), According to First Approximation (Formula [8]) and Second Approximation (Formula [11])**

| | Sample Size Necessary to Gain 80% Power (True Power of TDT[a]) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Multiplicative MOI | | Additive MOI | | Recessive MOI | | Dominant MOI | |
| $\gamma_{AND}$ $p$ | First Approximation | Second Approximation | First Approximation | Second Approximation | First Approximation | Second Approximation | First Approximation | Second Approximation |
| 4.0: | | | | | | | | |
| .01 | 230 (.80) | 245 (.86) | 251 (.80) | 266 (.85) | 724,763 (.80) | 724,971 (.80) | 258 (.80) | 273 (.85) |
| .10 | 48 (.81) | 50 (.85) | 76 (.81) | 78 (.83) | 1,121 (.80) | 1,133 (.81) | 95 (.81) | 96 (.82) |
| .50 | 61 (.82) | 63 (.84) | 132 (.80) | 135 (.82) | 94 (.80) | 96 (.82) | 492 (.80) | 494 (.80) |
| .80 | 158 (.81) | 168 (.86) | 359 (.80) | 367 (.82) | 175 (.80) | 184 (.84) | 7,193 (.80) | 7,203 (.80) |
| 2.0: | | | | | | | | |
| .01 | 1,954 (.80) | 1,997 (.82) | 1,954 (.80) | 1997 (.82) | 12,404,460 (.80) | 12,404,685 (.80) | 2,049 (.80) | 2,091 (.82) |
| .10 | 263 (.80) | 267 (.82) | 263 (.80) | 267 (.82) | 14,940 (.80) | 14,956 (.80) | 399 (.80) | 402 (.81) |
| .50 | 179 (.80) | 181 (.81) | 179 (.80) | 181 (.81) | 424 (.80) | 426 (.81) | 1,108 (.80) | 1,111 (.80) |
| .80 | 392 (.80) | 399 (.82) | 392 (.80) | 399 (.82) | 489 (.80) | 495 (.81) | 14,328 (.80) | 14,337 (.80) |
| 1.5: | | | | | | | | |
| .01 | 7,752 (.80) | 7,813 (.81) | 7,449 (.80) | 7,511 (.81) | 60,955,123 (.80) | 60,955,300 (.80) | 8,012 (.80) | 8,072 (.81) |
| .10 | 939 (.80) | 945 (.81) | 684 (.80) | 690 (.81) | 69,839 (.80) | 69,857 (.80) | 1,292 (.80) | 1,297 (.80) |
| .50 | 484 (.80) | 486 (.80) | 229 (.80) | 231 (.81) | 1,422 (.80) | 1,424 (.80) | 2,546 (.80) | 2,548 (.80) |
| .80 | 939 (.80) | 945 (.80) | 419 (.80) | 425 (.82) | 1,268 (.80) | 1,273 (.80) | 30,064 (.80) | 30,073 (.80) |

[a] Estimated by simulation with $10^5$ replicates.

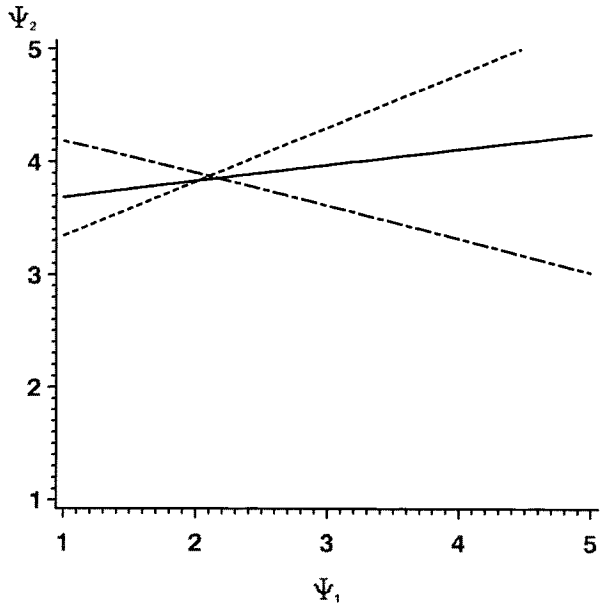


**Figure 1** Relative risks ($\Psi_1$, $\Psi_2$) resulting in a power of 90% for a sample of 95 nuclear families each with a single affected child ($\alpha = .05$) and different assumed population allele frequencies (unbroken line, $p = .37$; dotted line, $p = .42$; dashed-dotted line, $p = .32$).

low (up to a factor of 2) for $p < .5$ and too large for $p > .5$.

### *Sib Pairs*

The results of the simulation, for families each with an ASP, are presented in table 4. Again, the second approximation, formula (11), tends to overestimate the sample size necessary to gain 80% power, whereas the simulated power for sample sizes obtained by the first approximation, formula (8), agrees quite well with the expected power of .80. The sample sizes given by Camp (1997, table 3) are too low for $p < .5$ and too large for $p \geq .5$. For example, the sample size of 59 sib pairs, given by Camp (1997) for a multiplicative model with $\gamma = 4$ and $p = .01$ results in a power of <.002 (instead of .80).

### *A Real-Data Application*

Bellivier et al. (1998) have studied the association between bipolar I/II disorder and a biallelic polymorphism (A218C) of the tryptophan hydroxylase (TPH) gene, by means of a case-control study with 152 patients and 94 healthy control subjects. The frequency of the TPH *A* allele was .37 in their control group. The genotypic distributions for patients and controls were significantly different ($P = .002$). The odds ratio of bipolar disorder was 3.96 for homozygous *AA* individuals and 1.96 for heterozygous *AC* individuals, compared with homozygous *CC* individuals. M. Rietschel (personal communication) tried to replicate this finding by means of a family-based controls-association study. For this purpose, a sample of 95 nuclear families each consisting of a single affected offspring with bipolar I disorder plus both parents was available. The first power approximation described in the present article was used to evaluate the potential resolution of this intended replication study ($\alpha = .05$). The result is shown in figure 1. The unbroken line in figure 1 connects points ($\Psi_1$,$\Psi_2$), for which the power obtained with 95 nuclear families equals 90% under the assumption that the population frequency of allele *A* is .37. Thus, for points ($\Psi_1$,$\Psi_2$), which are above this unbroken line, the power is >90%. The dotted line and the dashed-dotted line in figure 1 give the analogous result when population frequencies .42 and .32, respectively, are assumed. As can be seen from figure 1, the study by Rietschel et al. would detect an allele-*A* effect, of the magnitude described by Bellivier et al. (1998), with >90% power, given that the population frequency of allele *A* is .32–.42.

## Discussion

The present article has applied standard statistical large-sample theory to approximate the power of the TDT for a general MOI of disease; thereby, it corrects sample-size formulas given by Camp (1997). Whereas it is always possible to estimate the power of the TDT against a specified alternative by simulations, the availability of a precise analytical power approximation is most useful for fast power calculations over a broad range of alternatives. For example, it would have been quite difficult and time-consuming to obtain, solely by means of simulations, the information presented in figure 1. Therefore, the availability of an accurate power approximation for the TDT will be valuable for the planning of a linkage and association study.

## Appendix

We have

$$\partial g/\partial x_1 = [x_1 + x_2 - \tfrac{1}{2} \cdot (x_1 - x_2)]/\left[(x_1 + x_2)^{\frac{3}{2}}\right],$$

$$\partial g/\partial x_2 = -[x_1 + x_2 + \tfrac{1}{2} \cdot (x_1 - x_2)]/\left[(x_1 + x_2)^{\frac{3}{2}}\right].$$

Therefore,

$$\begin{aligned}\sigma_{A1}^2 &= \{[e_1 + e_2 - \tfrac{1}{2}\cdot(e_1 - e_2)]^2/(e_1 + e_2)^3\}\cdot\sigma_{11}^* \\ &\quad -2\Big(\{[e_1 + e_2 - \tfrac{1}{2}\cdot(e_1 - e_2)] \\ &\quad \times[e_1 + e_2 + \tfrac{1}{2}\cdot(e_1 - e_2)]\}/(e_1 + e_2)^3\Big)\cdot\sigma_{12}^* \\ &\quad +\{[e_1 + e_2 + \tfrac{1}{2}\cdot(e_1 - e_2)]^2/(e_1 + e_2)^3\}\cdot\sigma_{22}^* \\ &= [(\sigma_{11}^* - 2\sigma_{12}^* + \sigma_{22}^*)/(e_1 + e_2)] \\ &\quad +\{[(e_1 - e_2)\cdot(\sigma_{22}^* - \sigma_{11}^*)]/(e_1 + e_2)^2\} \\ &\quad +\frac{1}{4}\cdot[(e_1 - e_2)^2\cdot(\sigma_{11}^* + 2\sigma_{12}^* + \sigma_{22}^*)]/(e_1 + e_2)^3\ ,\end{aligned}$$

which implies formula (7), in view of

$$\begin{aligned}\sigma_{11}^* - 2\sigma_{12}^* + \sigma_{22}^* &= d_1 - (e_1 - e_2)^2\ , \\ \sigma_{22}^* - \sigma_{11}^* &= d_{1,2} + (e_1 - e_2)\cdot(e_1 + e_2)\ , \\ \sigma_{11}^* + 2\sigma_{12}^* + \sigma_{22}^* &= d_2 - (e_1 + e_2)^2\ .\end{aligned}$$

## References


Baur MP, Knapp M (1997) Association studies in genetic epidemiology. In: Pawlowitzki IH, Edwards JH, Thompson EA (eds) Genetic mapping of disease genes. Academic Press, London, pp 159–172

Bellivier F, Leboyer M, Courtet P, Buresi C, Beaufils B, Samolyk D, Allilaire JF, et al (1998) Association between the tryptophan hydroxylase gene and manic-depressive illness. Arch Gen Psychiatry 55:33–37

Blackwelder WC, Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 2:85–97

Camp NJ (1997) Genomewide transmission/disequilibrium testing—consideration of the genotypic relative risks at disease loci Am J Hum Genet 61:1424–1430

Johnson NL, Kotz S (1972) Distributions in statistics: continuous multivariate distributions. John Wiley, New York

Lander ES (1996) The new genomics: global views of biology. Science 274:536–539

Rao CR (1972) Linear statistical inference and its applications. 2d ed. John Wiley, New York

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

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

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