Analytic power and sample size calculation for the genotypic transmission/disequilibrium test in case-parent trio studies

Christoph Neumann[1], **Margaret A. Taub**[2], **Samuel G. Younkin**[2], **Terri H. Beaty**[2], **Ingo Ruczinski**[2], and **Holger Schwender**[3,*]

[1]Faculty of Statistics, TU Dortmund University, 44221 Dortmund, Germany

[2]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21218, USA

[3]Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

Abstract

Case-parent trio studies considering genotype data from children affected by a disease and from their parents are frequently used to detect single nucleotide polymorphisms (SNPs) associated with disease. The most popular statistical tests in this study design are transmission/disequilibrium tests (TDTs). Several types of these tests have been developed, e.g., procedures based on alleles or genotypes. Therefore, it is of great interest to examine which of these tests have the highest statistical power to detect SNPs associated with disease. Comparisons of the allelic and the genotypic TDT for individual SNPs have so far been conducted based on simulation studies, since the test statistic of the genotypic TDT was determined numerically. Recently, it, however, has been shown that this test statistic can be presented in closed form. In this article, we employ this analytic solution to derive equations for calculating the statistical power and the required sample size for different types of the genotypic TDT. The power of this test is then compared with the one of the corresponding score test assuming the same mode of inheritance as well as the allelic TDT based on a multiplicative mode of inheritance, which is equivalent to the score test assuming an additive mode of inheritance. This is, thus, the first time that the power of these tests are compared based on equations, yielding instant results and omitting the need for time-consuming simulation studies. This comparison reveals that the tests have almost the same power, with the score test being slightly more powerful.

Keywords

Case-parent trio design; Conditional logisitc regression; Genome-wide association studies; Power calculation; Wald test

---

*Corresponding author: schwender@math.uni-duesseldorf.de. Phone: +49-211-81-15709, Fax: +49-211-81-10784.

1 Introduction

Case-parent trio studies are frequently used to test SNPs for association with disease by analyzing the genotypes of children having this disease and their parents. Advantages of case-parent trio and other family-based designs over population-based case-control studies are their robustness against spurious findings due to population stratification and the possibility to test for association and linkage simultaneously (Spielman and Ewens, 1996; Gauderman *et al*., 1999; Laird and Lange, 2006).

One of the most popular tests for association in case-parent trio studies is the allelic transmission/ disequilibrium test introduced by Spielman *et al*. (1993), which is equivalent to McNemar's test comparing the numbers of alleles transmitted or not transmitted from the heterozygous parents to their offspring affected by disease. This allelic TDT thus allows the detection of alleles preferentially transmitted to the affected offspring, and hence, potentially associated with the disease of the children.

Instead of testing alleles (and thus, considering chromosomes as units in the analysis), genotypes (and therefore, individuals) can also be directly analyzed by employing a genotypic transmission/disequilibrium test. In the genotypic TDT, the genotype of an affected child is compared to the three other genotypes possible given the parents' genotypes, but not shown by the affected offspring (Self *et al*., 1991; Schaid, 1996). This test is equivalent to a Wald test in a conditional logistic regression model in which each caseparent trio forms a stratum and the respective three not transmitted genotypes serve as controls (usually referred to as pseudo-controls, as these controls are artificial).

While the allelic TDT is based on the assumption of a multiplicative mode of inheritance, the genotypic TDT can be used to test a wide range of genetic models, considering, e.g., an additive, dominant, or recessive mode of inheritance (see, e.g., Fallin *et al*., 2002). Moreover, the genotypic TDT allows the determination of parameter estimates, relative risks, standard errors, and confidence intervals in addition to p-values. These estimates can, e.g., be used to combine results from different case-parent trio studies as well as in meta-analyses of case-parent trio with population-based case-control studies (see, e.g., Ludwig *et al*., 2012). By contrast, both the allelic TDT and the score test corresponding to the Wald test in the conditional logistic regression model only provide (scores and) p-values.

A disadvantage of the genotypic TDT, in particular in genome-wide association studies, over the allelic TDT and the score test was its high computation time, as the likelihood of the conditional logistic regression model had to be maximized by employing an iterative procedure to obtain the test statistic of the Wald test, and hence, the genotypic TDT statistic. However, it has recently been shown that when testing SNPs individually an analytic solution for the maximum-likelihood estimator in this model, and thus, for the genotypic TDT statistic exists, no matter whether an additive, dominant, or recessive mode of inheritance is assumed (Schwender *et al*., 2012). Therefore, this drawback has been eliminated, and genome-wide applications of the genotypic TDT are as fast as analyses with the allelic TDT or the score test (see, in particular, Table 4 in Schwender *et al*., 2012).

These closed-form solutions of the genotypic TDT also allow the analytic determination of the power and sample size of the genotypic TDT (and the score test) assuming different modes of inheritance. This hence avoids the need for time-consuming computations of required sample sizes or power based on simulation studies. In this article, we derive equations for these power and sample size determinations and compare the sample sizes required by the genotypic TDTs to reach a certain power with the ones needed by the corresponding score tests. This comparison also includes the allelic TDT proposed by Spielman *et al*. (1993) assuming a multiplicative mode of inheritance, since this test is equivalent to a score test assuming an additive mode of inheritance (cf. Schaid and Sommer, 1994).

For the allelic TDT under general modes of inheritance, equations for the approximations of power and sample sizes have already been devised by Knapp (1999). An alternative approach to the one of Knapp (1999) for power and sample size calculations for studies with a dichotomous outcome have been proposed by Lange and Laird (2002). Their procedure covers the wide range of general FBATs (Family-Based Association Tests) as suggested, e.g., by Laird *et al*. (2000) and Rabinowitz and Laird (2000) for different family-based designs and different situations in which, e.g., the genotypes of one or both parents are missing. This also includes the original TDT (i.e. the allelic TDT) proposed by Spielman *et al*. (1993). This method, however, does not cover the genotypic TDT, and hence, does not provide the possibility for analytic power calculation for the genotypic TDT. Moreover, a related approach has been devised by Lange *et al*. (2002) for power and sample size determinations for general FBATs considering quantitative traits.

This article is organized as follows: We first describe the analytic determination of the genotypic TDT statistics in Section 2. Afterwards, we derive in Sections 3 and 4 equations for power and sample size calculations, respectively, for the genotypic TDT. We focus in these sections on an additive mode of inheritance. Equations for power determinations for the dominant and recessive mode of inheritance as well as their derivation can be found in Appendix A.1. In Section 5, we furthermore present concise equations for the test statistics of the score test, assuming an additive mode of inheritance that can – analogously to the genotypic TDT statistics – be used for power and sample size calculation. Score tests for a dominant or a recessive mode of inheritance are discussed in Appendix A.2. In Section 6, the required sample sizes of these tests are compared with each other and with the ones of the allelic TDT determined based on the approach of Knapp (1999). Finally, we conduct a simulation study in Section 7 to validate the equations for the statistical power determination.

All the closed-form solutions for performing sample size and power calculation are implemented in the R-package trio freely available at http://www.bioconductor.org.

2 Analytic solution to the genotypic TDT

To test under a specified mode of inheritance (e.g., an additive, dominant, or recessive mode of inheritance) whether a SNP is associated with disease, the genotypic TDT assesses whether a genotype of this SNP is preferentially transmitted from the parents to their

affected offspring. The genotypic TDT is based on a conditional logistic regression model consisting only of one explanatory variable X coding for the specified mode of inheritance. In this model, the genotype of the affected child is compared with the three not transmitted genotypes that would have also been possible given the genotypes of the parents.

As an example assume that at a specific SNP one parent shows the homozygous reference genotype A_1A_2, and the other the heterozygous genotype A_3G (where the indices are only used to differ between the different alleles). Since each parent transmits one of these alleles to their offspring, it will exhibit one of the genotypes A_1A_3, A_1G, A_2A_3, and A_2G (cf. first two columns of Table 1). In the conditional logistic regression model, the genotype of this offspring is considered as case and the other three not transmitted genotypes are used as pseudo-controls, where the dependency structure is taken into account by forming one stratum for each case-parent trio.

Denoting the value of X for the affected offspring in case-parent trio $i = 1, \ldots, n$, by x_{i0}, and the values for the corresponding three pseudo-controls by x_{ik}, $k = 1, \ldots, 3$, the maximum-likelihood estimate for the parameter γ corresponding to X is determined by the value $\hat{\gamma}$ maximizing the conditional likelihood

$$L(\gamma) = \prod_{i=1}^{n} \frac{\exp(\gamma x_{i0})}{\sum_{k=0}^{3} \exp(\gamma x_{ik})} \quad (1)$$

of the conditional logistic regression model. The test statistic of the genotypic TDT is then given by the Wald statistic

$$g^2 = \frac{\hat{\gamma}^2}{\mathrm{Var}\,(\hat{\gamma})}. \quad (2)$$

The likelihood (1) has to be maximized over the n weights $w_i = \exp(\gamma x_{i0}) / \sum_{k=0}^{3} \exp(\gamma x_{ik})$ of the n trios. Considering the above example trio and assuming that the offspring shows one of the heterozygous genotypes, the weight of this case-parent trio under an additive mode of inheritance (in which case, X codes for the number of minor alleles) is given by

$$w(\gamma_{\mathrm{add}}) = \frac{\exp(\gamma_{\mathrm{add}} \cdot 1)}{2\exp(\gamma_{\mathrm{add}} \cdot 0) + 2\exp(\gamma_{\mathrm{add}} \cdot 1)} = \frac{\exp(\gamma_{\mathrm{add}})}{2 + 2\exp(\gamma_{\mathrm{add}})}.$$

However, there only exist ten possible genotype combinations for case-parent trios, and thus, (at most) ten different weights (see Table 1). Since three of these genotype combinations have weights not depending on γ_{add}, only seven of them – namely the ones comprising at least one heterozygous parent – contribute to the maximization of (1). In this situation, the logarithm of the conditional likelihood (1), therefore, reduces to a sum over seven numbers $n_c^{(p_1,p_2)}$ of trios showing the respective genotype combination weighted by $w(\gamma_{\mathrm{add}})$, where $c, p_1, p_2 \in \{0, 1, 2\}$ with $p_1 \quad p_2$ are the numbers of minor alleles of the

children and their parents in the respective trios. Using these numbers and the weights from Table 1, the reduced log-likelihood is thus given by

$$
\begin{aligned}
\ell^*(\gamma_{\mathrm{add}}) = & \left(n_0^{(0,1)} + n_1^{(1,2)} \right) \log \left(\frac{1}{2+2\exp(\gamma_{\mathrm{add}})} \right) \\
& + \left(n_1^{(0,1)} + n_2^{(1,2)} \right) \log \left(\frac{\exp(\gamma_{\mathrm{add}})}{2+2\exp(\gamma_{\mathrm{add}})} \right) \\
& + \sum_{c=0}^{2} \log \left(\frac{\exp(c\gamma_{\mathrm{add}})}{(1+\exp(\gamma_{\mathrm{add}}))^2} \right) n_c^{(1,1)} \\
= & -\left(\log(2)\right. \\
& + \log(1 \\
& \left. + \exp(\gamma_{\mathrm{add}})\right) \left(n_0^{(0,1)} \right. \\
& \left. + n_1^{(0,1)} + n_1^{(1,2)} + n_2^{(1,2)} \right) + \left(n_1^{(0,1)} + n_2^{(1,2)} + n_1^{(1,1)} + 2n_2^{(2,2)} \right) \gamma_{\mathrm{add}} \\
& - 2\log(1 \\
& + \exp(\gamma_{\mathrm{add}})) \sum_{c=0}^{2} n_c^{(1,1)}.
\end{aligned}
\tag{3}
$$

Noticing that

$$
n_{\mathrm{het}} = n_0^{(0,1)} + n_1^{(0,1)} + n_1^{(1,2)} + n_2^{(1,2)} + 2\sum_{c=0}^{2} n_c^{(1,1)} \tag{4}
$$

is the total number of heterozygous parents and

$$
n_{\mathrm{not}} = n_1^{(0,1)} + n_2^{(1,2)} + n_1^{(1,1)} + 2n_2^{(1,1)} \tag{5}
$$

is the total number of more frequent alleles not transmitted from the heterozygous parents to their affected offspring – or analogously, the total number of minor alleles transmitted by the heterozygous parents – the first derivative of (3) is given by

$$
\frac{\partial \ell^*(\gamma_{\mathrm{add}})}{\partial \gamma_{\mathrm{add}}} = n_{\mathrm{not}} - \frac{\exp(\gamma_{\mathrm{add}})}{1+\exp(\gamma_{\mathrm{add}})} n_{\mathrm{het}}. \tag{6}
$$

Setting (6) to zero and solving it for γ_{add}, the maximum-likelihood estimator for γ_{add} is given by

$$
\hat{\gamma}_{\mathrm{add}} = \mathrm{logit} \left(\frac{n_{\mathrm{not}}}{n_{\mathrm{het}}} \right) = \log \left(\frac{n_{\mathrm{not}}}{n_{\mathrm{het}} - n_{\mathrm{not}}} \right). \tag{7}
$$

The variance of $\hat{\gamma}_{\mathrm{add}}$ can then be estimated by the value of the negative inverse of the second derivative

$$\frac{\partial^2 \ell(\gamma_{\text{add}})}{\partial \gamma_{\text{add}}^2} = -\frac{\exp(\gamma_{\text{add}})}{(1+\exp(\gamma_{\text{add}}))^2} n_{\text{het}} \quad (8)$$

at $\gamma_{\text{add}} = \hat{\gamma}_{\text{add}}$, i.e. by

$$\hat{\text{Var}}\left(\hat{\gamma}_{\text{add}}\right) = \frac{n_{\text{het}}}{(n_{\text{het}} - n_{\text{not}}) n_{\text{not}}},$$

For a more detailed discussion of the analytic solution for the genotypic TDT, see Schwender *et al.* (2012).

Analogously, closed-form solutions for the genotypic TDT statistic (2) can be derived for the dominant and the recessive mode of inheritance (for these derivations, see Schwender *et al.*, 2012), where in the dominant case the coding variable *X* is set to 0 if the subject shows the homozygous reference genotype, and to 1 otherwise. For a recessive mode of inheritance, *X* is set to 1 if both chromosomes show the minor allele, and to 0 otherwise. In both cases, the maximum likelihood estimate for γ takes the form

$$\hat{\gamma} = \log\left(\sqrt{a+h^2} - h\right),$$

where, e.g., in a dominant model *a* and *h* are given by

$$a_{\text{dom}} = \frac{n_1^{(0,1)} + n_1^{(1,1)} + n_2^{(1,1)}}{3\left(n_0^{(0,1)} + n_0^{(1,1)}\right)} \quad (9)$$

and

$$h_{\text{dom}} = \left(\frac{1/3\left(n_0^{(0,1)} - n_1^{(1,1)} - n_2^{(1,1)}\right) - n_1^{(0,1)} + n_0^{(1,1)}}{2\left(n_0^{(0,1)} + n_0^{(1,1)}\right)}\right), \quad (10)$$

respectively.

3 Power calculation for the genotypic TDT

For the additive mode of inheritance, the power of the genotypic TDT can be determined by an approach analogous to the one used by Knapp (1999) for calculating the power of the allelic TDT. For this, we consider the random vector $\mathbf{Z}_{\text{het}} = (Z_1, \ldots, Z_7)^T$ consisting of random variables $Z_j, j = 1, \ldots, 7$, for the seven numbers of trios corresponding to the genotype combinations that influence the maximization of the log-likelihood (3). This random vector is a subvector of $\mathbf{Z} = (Z_1, \ldots, Z_8)^T$, where $\mathbf{Z}$ additionally contains the random variable Z_8 specifying the total number of trios belonging to the other three genotype combinations without heterozygous parents. This random vector $\mathbf{Z}$ is thus multinomially distributed with n observations (here, trios) and probability vector $\mathbf{q} = (q_1, \ldots, q_8)^T$.

We further define $\mathbf{u} = (u_1, \ldots, u_7)^T$ and $\mathbf{v} = (v_1, \ldots, v_7)^T$ as the vectors containing the numbers u_j and v_j of more frequent alleles transmitted and not transmitted, respectively, from the heterozygous parents to their offspring in trios with the j-th genotype combination. Using these specifications, it can be derived from (4) and (5) that

$$N_{\text{het}} = \mathbf{u}^T \mathbf{Z}_{\text{het}} + \mathbf{v}^T \mathbf{Z}_{\text{het}} \text{ and } N_{\text{not}} = \mathbf{v}^T \mathbf{Z}_{\text{het}} \quad (11)$$

are the random variables generating n_{het} and n_{not}, respectively. Therefore, the square root of the test statistic of the genotypic TDT can be rewritten as

$$G = \log\left(\frac{V}{U}\right) \sqrt{\frac{UV}{U+V}}$$

with $U = \mathbf{u}^T \mathbf{Z}_{\text{het}}/n$ and $V = \mathbf{v}^T \mathbf{Z}_{\text{het}}/n$. Note that under the null hypothesis $H_0 : \gamma = 0$ G is standard normally distributed so that G^2 is χ^2-distributed with one degree of freedom.

Following the same arguments as Knapp (1999) based on the theoretic results presented in Rao (1973) and setting $\tilde{u} = \mathbf{u}^T \mathbf{q}_{\text{het}}$ and $\tilde{v} = \mathbf{v}^T \mathbf{q}_{\text{het}}$ with $\mathbf{q}_{\text{het}} = (q_1, \ldots, q_7)^T$, the test statistic G follows asymptotically a normal distribution with mean

$$\mu_{\text{add}} = \sqrt{n} \mathrm{E}_1(g_{\text{add}}) = \sqrt{n}\log\left(\frac{\tilde{u}}{\tilde{v}}\right) \sqrt{\frac{\tilde{u}\tilde{v}}{\tilde{u}+\tilde{v}}}$$

and variance

$$\sigma^2_{\text{add}} = \mathbf{u}^T \mathbf{\Sigma} \mathbf{u} \left(\frac{\partial g}{\partial \tilde{u}}\right)^2 + 2\mathbf{u}^T \mathbf{\Sigma} \mathbf{v} \frac{\partial g}{\partial \tilde{u}} \frac{\partial g}{\partial \tilde{v}} + \mathbf{v}^T \mathbf{\Sigma} \mathbf{v} \left(\frac{\partial g}{\partial \tilde{v}}\right)^2 .$$

Here, $\mathrm{E}_1(g_{\text{add}})$ is the expected value of the genotypic TDT statistic for $n = 1$, $\mathbf{\Sigma}$ is a 7×7 matrix with diagonal elements $q_j(1 - q_j)$ and off-diagonal elements $-q_j q_\ell$, and the two derivatives are given by

$$\frac{\partial g}{\partial \tilde{u}} = \frac{\left(\frac{1}{2}\log\left(\frac{\tilde{v}}{\tilde{u}}\right) - 1\right)\sqrt{(\tilde{u}+\tilde{v})\frac{\tilde{v}}{\tilde{u}}} - \frac{1}{2}\log\left(\frac{\tilde{v}}{\tilde{u}}\right)\sqrt{\frac{\tilde{u}\tilde{v}}{\tilde{u}+\tilde{v}}}}{\tilde{u}+\tilde{v}} \quad \frac{\partial g}{\partial \tilde{v}} = \frac{\left(\frac{1}{2}\log\left(\frac{\tilde{v}}{\tilde{u}}\right) + 1\right)\sqrt{(\tilde{u}+\tilde{v})\frac{\tilde{u}}{\tilde{v}}} - \frac{1}{2}\log\left(\frac{\tilde{v}}{\tilde{u}}\right)\sqrt{\frac{\tilde{u}\tilde{v}}{\tilde{u}+\tilde{v}}}}{\tilde{u}+\tilde{v}} .$$

The statistical power β_{add} of the genotypic TDT assuming an additive mode of inheritance is thus asymptotically given by

$$\beta_{\text{add}} = \Phi\left(\frac{z_{\alpha/2} - \mu_{\text{add}}}{\sigma_{\text{add}}}\right) + 1 - \Phi\left(\frac{z_{1-\alpha/2} - \mu_{\text{add}}}{\sigma_{\text{add}}}\right), \quad (12)$$

where Φ is the cumulative distribution function of the standard normal distribution and z_α is the α-quantile of the standard normal distribution.

In a case-parent trio study, the values $q_j, j = 1, \ldots, 7$, can be estimated, and thus, the power of the genotypic TDT can be determined for each SNP from the data. It is, however, often also of interest to compute the power for a given number n of trios, a type I error rate α, and a relative risk RR. Since under the assumption of Hardy-Weinberg equilibrium in the parents there exists a direct relationship between the relative risk and the probabilities q_j for the different types of trios, $q_j, j = 1, \ldots, 7$, can be computed from the relative risk RR (see Schaid, 1999, for general equations for these probabilities).

For such a determination in an additive model, we set $r_0 = 1$, $r_1 = RR$, and $r_2 = 2RR - 1$, where r_c is the risk to get the disease with c minor alleles relative to the disease risk with no minor allele (Schaid, 1999). Further denoting the minor allele frequency by m, the probabilities for the different types of trios can be computed by

$$q_j = \frac{2m^{4-p_1-p_2}(1-m)^{p_1+p_2}r_c}{\max\{u_j, v_j\} \cdot (2m(RR-1)+1)},$$

where p_1 and p_2 are the numbers of minor alleles of the parents (as defined in Section 2).

Equations for the statistical power of the genotypic TDT assuming either a dominant or recessive mode of inheritance can be devised in a similar way as for the additive model. For the derivation of these equations, see Appendix A.1.

4 Sample size calculation for the genotypic TDT

An equation for the required sample size for a given type I error rate α and power β can be derived from equation (12) for the statistical power in the standard way. If α is small and the relative risk is not too close to 1, either $\Phi\left(\frac{z_{\alpha/2} - \mu_{\mathrm{add}}}{\sigma_{\mathrm{add}}}\right)$ or $1 - \Phi\left(\frac{z_{1-\alpha/2} - \mu_{\mathrm{add}}}{\sigma_{\mathrm{add}}}\right)$ becomes virtually zero so that this term of (12) does only very slightly influence the statistical power. Due to the symmetry of these terms, the sample size n required to gain a desired power β and to control the type I error rate at α can in both situations be determined by

$$n \approx \left(\frac{\sigma z_\beta + z_{1-\alpha/2}}{\mathrm{E}_1(g)}\right)^2. \quad (13)$$

5 Power and sample size determination for the score tests

The test statistic of a score test for testing the null hypothesis $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma \ 0$ is given by

$$s_{\mathrm{add}}^2 = \frac{D^2(0)}{I(0)}$$

with

$$D(\gamma) = \frac{\partial \ell(\gamma)}{\partial \gamma} \text{ and } I(\gamma) = -\frac{\partial^2 \ell(\gamma)}{\partial \gamma^2}.$$

Assuming an additive mode of inheritance, these two derivatives are given by (7) and (8) so that the test statistic assuming an additive genetic model can be determined by

$$s^2_{\text{add}} = \frac{(n_{\text{not}} - 0.5 n_{\text{het}})^2}{0.25 n_{\text{het}}} = \frac{(2 n_{\text{not}} - n_{\text{het}})^2}{n_{\text{het}}}. \quad (14)$$

As shown by Schaid and Sommer (1994), this test statistic is equivalent to the test statistic of the allelic TDT (Spielman *et al.*, 1993). Using our notation, this can be shown by noting that $n_{\text{het}} = n_A + n_a$ and $n_{\text{not}} = n_A$, where n_A and n_a are the numbers of the minor and the more frequent allele, respectively, transmitted from heterozygous parents to their offspring (cf. (11)). Therefore, (14) becomes

$$s^2_{\text{add}} = \frac{(2 n_A - (n_A + n_a))^2}{n_A + n_a} = \frac{(n_A - n_a)^2}{n_A + n_a},$$

which is the test statistic of McNemar's test, i.e. the allelic TDT. Power and required sample size of the score test assuming an additive mode of inheritance can hence be determined by exactly the same approach proposed for the allelic TDT assuming a multiplicative model by Knapp (1999).

For a discussion of the score tests assuming either a dominant or recessive mode of inheritance and the power calculation for these tests, see Appendix A.2.

6 Comparison of genotypic TDT and score test

Based on the approaches presented in the previous sections the sample sizes required by the genotypic TDTs to gain a certain power can be compared with the required sample sizes of the corresponding score tests.

Using equation (13), we thus computed the required sample sizes for these tests assuming an additive, dominant, or recessive mode of inheritance, considering different values of the relative risk (RR = 1.05, 1.20, 1.30, 1.40, 1.50) lying in the range of the relative risks observed in association studies as well as different minor allele frequencies (MAF = 0.01, 0.10, 0.20, 0.50) also considered by Knapp (1999) and Schaid (1999). As type I error rate we chose $\alpha = 5 \times 10^{-8}$, which is often used to call (genome-wide) significance in genome-wide association studies. Moreover, we considered a power of $\beta = 0.80$ often desired to be gained in a study.

The results of this comparison are summarized in Table 2. This table reveals that the sample size required by the score test is always slightly smaller than the one needed by the corresponding genotypic TDT. Compared to the total sample sizes, these differences are, however, virtually negligible. This table also supports the well-known fact that huge

numbers of subjects are required to detect with a high power genome-wide significant SNPs with a realistic relative risk.

We also compared the sample sizes required by the genotypic TDTs and the score tests with the ones determined by the approach proposed by Knapp (1999) for approximating the power of the allelic TDT for general modes of inheritance. The latter sample sizes were previously published in Table 3 of Knapp (1999). For this comparison, we computed the sample sizes required by the genotypic TDT and the score test using the same settings considered in Knapp (1999). These sample sizes are summarized in Table 3.

In particular in the recessive model, the required sample sizes for both the genotypic TDT and the score test are much smaller than ones computed for the allelic TDT based on the approach of Knapp (1999).

7 Simulation study

To validate the accuracy of the proposed sample size calculation, we performed a simulation study using the minor allele frequencies and the type I error rate considered in the previous section. Since the expected value of the estimate for the parameter γ in the conditional logistic regression model on which the genotypic TDT is based is the log relative risk (Schaid, 1996), we considered 1.05, 1.20, 1.30, 1.40, and 1.50 as values for $\exp(\gamma)$. For each combination of these minor allele frequencies and $\exp(\gamma)$, we simulated 10^5 case-parent trio data sets of the respective sample size determined in the previous section. The power was then estimated by the proportion of data sets for which the null hypothesis was rejected at the $\alpha = 5 \times 10^{-8}$ level. Because of the huge sample sizes for the recessive model, we only considered genotypic TDTs and score tests assuming an additive or dominant mode of inheritance.

The results of this simulation study are summarized in Table 4, which shows that all estimated powers are very close to $\beta = 0.80$, the power used in Section 6. Therefore, the sample size and power equations derived in Sections 3–5 as well as the Appendix show up to be very accurate even for relative risks close to 1 and small minor allele frequencies.

8 Discussion

We have presented equations for determining the statistical powers and the required sample sizes for the genotypic TDT and the corresponding score test, assuming an additive, dominant, or recessive mode of inheritance. These approaches allow the determination of the power of the genotypic TDT for several relative risks, minor allele frequencies, etc., in less than a split second, and therefore, avoid very time-consuming simulation-based power and sample size estimations.

A comparison of the genotypic TDTs with the corresponding score tests showed that both require about the same sample size to gain the same power with a very slight advance for the score test. This comparison also implicitly contained the original allelic TDT, as this test (assuming a multiplicative mode of inheritance) is equivalent to the score test for an additive model.

This comparison also reconfirmed the well-known fact that a huge number of samples is required to gain an acceptable power to detect genome-wide significant SNPs with typically small relative risks, where the required sample size rapidly increases with decreasing minor allele frequency. One of the reasons for this is that the smaller the minor allele frequency, the less trios contribute to the maximization of the conditional likelihood considered when performing the genotypic TDT, or more generally, when computing the Wald statistic.

The power and sample size determinations presented in this paper are implemented in the R package trio version 3.1.2 or later freely available at http://www.bioconductor.org.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

A.1. Determination of the asymptotic normal distributions of genotypic TDT statistics

Equations for the statistical power of the genotypic TDT assuming either a dominant or recessive mode of inheritance can be derived in a similar way as for the additive model described in Section 3. In these cases, however, the number of components forming the genotypic TDT statistic cannot be reduced to two terms U and V as in the additive case, but four components have to be considered (e.g., in the dominant model presented in the Section 2, just $n_1^{(1,1)}$ and $n_2^{(1,1)}$ can be combined). Denoting these components by $b_1, \ldots, b_4$, the test statistic G of the genotypic TDT assuming either a dominant or recessive mode of inheritance also follows asymptotically a normal distribution with mean

$$\mu = \sqrt{n} E_1(g)$$

and variance

$$\sigma^2 = \sum_{i=1}^{4} \sum_{k=1}^{4} \sigma_{ik} \frac{\partial g}{\partial b_i} \frac{\partial g}{\partial b_k}, \quad (15)$$

where σ_{ik} are the pairwise variances of these terms computed as described in Section 3 (see also Knapp, 1999).

In the following, the derivatives $\frac{\partial g}{\partial b_i}$, $i = 1, \ldots, 4$, are determined for the dominant and the recessive mode of inheritance. To differ between these two cases, we use in the following

the notation d_i (instead of b_i) for the numbers of trios when considering a dominant mode of inheritance and the notation r_i in the recessive case (for the specification of these numbers, see Table 5).

A.1.1. Genotypic TDT assuming a dominant model

For the dominant mode of inheritance, the numerator and denominator of the genotypic TDT statistic $g_{\text{dom}}^2 = \hat{\gamma}_{\text{dom}}^2 / \text{Var}(\hat{\gamma}_{\text{dom}})$ are determined from

$$\hat{\gamma}_{\text{dom}} = \log\left(\sqrt{a_{\text{dom}} + h_{\text{dom}}^2} - h_{\text{dom}}\right) \quad (16)$$

with a_{dom} and h_{dom} as specified by (9) and (10), respectively, as well as

$$V_{\text{dom}} = \text{Var}^{-1}(\hat{\gamma}_{\text{dom}}) = \frac{\left(n_0^{(0,1)} + n_1^{(0,1)}\right)\exp(\hat{\gamma}_{\text{dom}})}{(\exp(\hat{\gamma}_{\text{dom}}) + 1)^2} + \frac{\sum_{c=1}^3 n_c^{(1,1)}\exp(\hat{\gamma}_{\text{dom}})}{3(\exp(\hat{\gamma}_{\text{dom}}) + 1/3)^2}.$$

The square root of the genotypic TDT statistic can, thus, be written as $g_{\text{dom}} = \hat{\gamma}_{\text{dom}}\sqrt{V_{\text{dom}}}$, and the first derivatives of g_{dom} with respect to d_i, $i = 1, \ldots, 4$, can, hence, be determined by

$$\frac{\partial g_{\text{dom}}}{\partial d_i} = \frac{\partial \hat{\gamma}_{\text{dom}}}{\partial d_i} \cdot \sqrt{V_{\text{dom}}} + \frac{\partial V_{\text{dom}}}{\partial d_i} \cdot \frac{\hat{\gamma}_{\text{dom}}}{2\sqrt{V_{\text{dom}}}}.$$

To devise the variance of the asymptotic normal distribution, we, therefore, need to compute $\frac{\partial \hat{\gamma}_{\text{dom}}}{\partial d_i}$ as well as $\frac{\partial V_{\text{dom}}}{\partial d_i}$, where in the dominant model the components $d_1, \ldots, d_4$ are given by

$$d_1 = n_0^{(0,1)}, d_2 = n_1^{(0,1)}, d_3 = n_0^{(1,1)}, d_4 = n_1^{(1,1)} + n_2^{(1,1)}$$

(see Table 5).

Differentiating (16) with respect to d_i, $i = 1, \ldots, 4$, leads to

$$\frac{\partial \hat{\gamma}_{\text{dom}}}{\partial d_i} = \frac{1}{\exp(\hat{\gamma}_{\text{dom}})}\left(\frac{a_{\text{dom}}^{(i)} + 2h_{\text{dom}}h_{\text{dom}}^{(i)}}{2\sqrt{a_{\text{dom}} + h_{\text{dom}}^2}} - h_{\text{dom}}^{(i)}\right), \quad (17)$$

where

$$h_{\text{dom}}^{(i)} = \frac{\partial h_{\text{dom}}}{\partial d_i} \quad \text{and} \quad a_{\text{dom}}^{(i)} = \frac{\partial a_{\text{dom}}}{\partial d_i}.$$

More exactly,

$$h_{\text{dom}}^{(1)} = \frac{3d_2 - 2d_3 + d_4}{6(d_1 + d_3)^2}, \; h_{\text{dom}}^{(2)} = -\frac{1}{2(d_1 + d_3)}, \; h_{\text{dom}}^{(3)} = \frac{2d_1 + 3d_2 + d_4}{6(d_1 + d_3)^2}, \; h_{\text{dom}}^{(4)} = -\frac{1}{6(d_1 + d_3)},$$

and

$$a_{\text{dom}}^{(1)} = a_{\text{dom}}^{(3)} = -\frac{d_2 + d_4}{3(d_1 + d_3)^2}, \; a_{\text{dom}}^{(2)} = a_{\text{dom}}^{(4)} = \frac{1}{3(d_1 + d_3)}.$$

Setting $c_1 = c_2 = 1$ and $c_3 = c_4 = 1/3$ as well as

$$t_{\text{dom}}^{(i)} = \frac{\partial}{\partial d_i} \sqrt{a_{\text{dom}} + h_{\text{dom}}^2} = \frac{a_{\text{dom}}^{(i)} + 2h_{\text{dom}} h_{\text{dom}}^{(i)}}{2\sqrt{a_{\text{dom}} + h_{\text{dom}}^2}},$$

the first derivative of V_{dom} with respect to d_i, $i = 1, \dots, 4$, can then be derived as

$$\frac{\partial V_{\text{dom}}}{\partial d_i} = \frac{c_i \exp(\hat{\gamma}_{\text{dom}})}{(\exp(\hat{\gamma}_{\text{dom}}) + c_i)^2} + \left(t_{\text{dom}}^{(i)} - h_{\text{dom}}^{(i)}\right) \sum_{k=1}^{4} \frac{c_k d_k (c_k - \exp(\hat{\gamma}_{\text{dom}}))}{(\exp(\hat{\gamma}_{\text{dom}}) + c_k)^3}.$$

A.1.2. Genotypic TDT assuming a recessive model

Under the assumption of a recessive mode of inheritance, the maxixmum likelihood estimator of γ_{rec} is given by

$$\hat{\gamma}_{\text{rec}} = \log\left(\sqrt{a_{\text{rec}} + h_{\text{rec}}^2} - h_{\text{rec}}\right), \quad (18)$$

where

$$a_{\text{rec}} = \frac{3(r_2 + r_4)}{r_1 + r_3} \text{ and } h_{\text{rec}} = \frac{3r_1 - r_2 + r_3 - 3r_4}{2(r_1 + r_3)}$$

contain the four components

$$r_1 = n_1^{(1,2)}, \; r_2 = n_2^{(1,2)}, \; r_3 = n_0^{(1,1)} + n_1^{(1,1)}, \; r_4 = n_2^{(1,1)}$$

(see also Table 5). Further, the inverse V_{rec} of the variance of (18) can be determined by

$$V_{\text{rec}} = \text{Var}^{-1}(\hat{\gamma}_{\text{rec}}) = (r_1 + r_2)\frac{\exp(\hat{\gamma}_{\text{rec}})}{(\exp(\hat{\gamma}_{\text{rec}}) + 1)^2} + 3(r_3 + r_4)\frac{\exp(\hat{\gamma}_{\text{rec}})}{(\exp(\hat{\gamma}_{\text{rec}}) + 3)^2} \quad (19)$$

Analogously to the dominant case, the variance of the asymptotic normal distribution can be derived by computing the first derivatives of $\hat{\gamma}_{\text{rec}}$ and V with respect to r_i, $i = 1, \dots, 4$.

Since $\hat{\gamma_{\mathrm{rec}}}$ has the same form as $\hat{\gamma_{\mathrm{dom}}}$, its first derivatives are identical to (17), except that a_{dom} and h_{dom} are replaced by a_{rec} and h_{rec}, respectively. These first derivatives are thus given by

$$\frac{\partial \hat{\gamma}_{\mathrm{rec}}}{\partial r_i} = \frac{1}{\exp(\hat{\gamma}_{\mathrm{rec}})} \left(\frac{a_{\mathrm{rec}}^{(i)} + 2 h_{\mathrm{rec}} h_{\mathrm{rec}}^{(i)}}{2 \sqrt{a_{\mathrm{rec}} + h_{\mathrm{rec}}^2}} - h_{\mathrm{rec}}^{(i)} \right)$$

with

$$h_{\mathrm{rec}}^{(1)} = \frac{r_2 + 2r_3 + 3r_4}{2(r_1 + r_3)^2}, \; h_{\mathrm{rec}}^{(2)} = -\frac{1}{2(r_1 + r_3)}, \; h_{\mathrm{rec}}^{(3)} = \frac{-2r_1 + r_2 + 3r_4}{2(r_1 + r_3)^2}, \; h_{\mathrm{rec}}^{(4)} = -\frac{3}{2(r_1 + r_3)},$$

and

$$a_{\mathrm{rec}}^{(1)} = a_{\mathrm{rec}}^{(3)} = -\frac{3(r_2 + r_4)}{(r_1 + r_3)^2}, \; a_{\mathrm{rec}}^{(2)} = a_{\mathrm{rec}}^{(4)} = \frac{3}{r_1 + r_3}.$$

The same applies to the first derivatives of (19), which take the same form as the first derivatives of V_{dom} in the dominant model. Setting $c_1 = c_2 = 1$, $c_3 = c_4 = 3$, and again

$$t_{\mathrm{rec}}^{(i)} = \frac{\partial}{\partial r_i} \sqrt{a_{\mathrm{rec}} + h_{\mathrm{rec}}^2} = \frac{a_{\mathrm{rec}}^{(i)} + 2 h_{\mathrm{rec}} h_{\mathrm{rec}}^{(i)}}{2 \sqrt{a_{\mathrm{rec}} + h_{\mathrm{rec}}^2}},$$

the first derivative of (19) with respect to r_i, $i = 1, \ldots, 4$, is, thus, given by

$$\frac{\partial V_{\mathrm{dom}}}{\partial r_i} = \frac{c_i \exp(\hat{\gamma}_{\mathrm{rec}})}{(\exp(\hat{\gamma}_{\mathrm{rec}}) + c_i)^2} + \left(t_{\mathrm{rec}}^{(i)} - h_{\mathrm{rec}}^{(i)} \right) \sum_{k=1}^{4} \frac{c_k r_k \left(c_k - \exp(\hat{\gamma}_{\mathrm{rec}}) \right)}{(\exp(\hat{\gamma}_{\mathrm{rec}}) + c_k)^3}.$$

A.2. Determination of the asymptotic normal distributions of score test statistics

The score test statistic is more complex when considering a dominant or a recessive mode of inheritance than when assuming an additive mode of inheritance. In the dominant case, the score test statistic is given by

$$s_{\mathrm{dom}}^2 = \frac{\left(2 \left(n_1^{(0,1)} - n_0^{(0,1)} \right) - 3 n_0^{(11)} + n_1^{(11)} + n_2^{(11)} \right)^2}{4 \sum_{c=1}^{2} n_c^{(0,1)} + 3 \sum_{c=0}^{2} n_c^{(1,1)}}.$$

For an recessive model, this test statistic can be determined by

$$s_{\mathrm{rec}}^2 = \frac{\left(2\left(n_2^{(1,2)} - n_1^{(1,2)}\right) - n_0^{(1,1)} - n_1^{(1,1)} + 3n_2^{(1,1)}\right)^2}{4\sum_{c=1}^{2} n_c^{(1,2)} + 3\sum_{c=0}^{2} n_c^{(1,1)}}$$

(for an alternative representation of these test statistics, see Schaid and Sommer, 1994).

Since these test statistics are concise, they can directly be differentiate with respect to the respective four components also considered in Appendix A.1 to derive the asymptotic normal distribution in the genotypic TDT. Setting

$$e_{\mathrm{dom}} = (4d_1 + 4d_2 + 3d_3 + 3d_4)^{3/2} \quad \text{and} \quad e_{\mathrm{rec}} = (4r_1 + 4r_2 + 3r_3 + 3r_4)^{3/2}$$

the first derivatives of s_{dom} with respect to $d_1, \ldots, d_4$ are given by

$$\frac{\partial s_{\mathrm{dom}}}{\partial d_1} = -\frac{4(d_1 + 3d_2 + 2d_4)}{e_{\mathrm{dom}}}, \qquad \frac{\partial s_{\mathrm{dom}}}{\partial d_2} = -\frac{4(3d_1 + d_2 + 3d_3 + d_4)}{e_{\mathrm{dom}}},$$
$$\frac{\partial s_{\mathrm{dom}}}{\partial d_3} = -\frac{2(6d_1 + 10d_2 + 3d_3 + 7d_4)}{3e_{\mathrm{dom}}}, \qquad \frac{\partial s_{\mathrm{dom}}}{\partial d_4} = -\frac{14d_1 + d_2 + 15d_3 + 3d_4}{2e_{\mathrm{dom}}},$$

and the first derivatives of s_{rec} with respect to $r_1, \ldots, r_4$ by

$$\frac{\partial s_{\mathrm{rec}}}{\partial r_1} = -\frac{4(r_1 + 3r_2 + r_3 + 3r_4)}{e_{\mathrm{rec}}}, \qquad \frac{\partial s_{\mathrm{rec}}}{\partial r_2} = \frac{4(3r_1 + r_2 + 2r_3)}{e_{\mathrm{rec}}},$$
$$\frac{\partial s_{\mathrm{rec}}}{\partial r_3} = -\frac{2r_1 + 14r_2 + 3r_3 + 15r_4}{2e_{\mathrm{rec}}}, \qquad \frac{\partial s_{\mathrm{rec}}}{\partial r_4} = \frac{2(10r_1 + 6r_2 + 7r_3 + 3r_4)}{3e_{\mathrm{rec}}}.$$

These derivatives can then be inserted into equation (15) to compute the variance of the asymptotic normal distribution of the square root S of the score test statistic.

References

Fallin D, Beaty T, Liang K-Y, Chen W. Power comparisons for genotypic vs. allelic TDT methods with > 2 alleles. Genetic Epidemiology. 2002; 23:458–461. [PubMed: 12432512]

Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Journal of the National Cancer Institute Monographs. 1999; 26:31–37. [PubMed: 10854483]

Knapp M. A note on power approximations for the transmission/disequilibrium test. American Journal of Human Genetics. 1999; 64:1177–1185. [PubMed: 10090903]

Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. Nature Reviews Genetics. 2006; 7:385–394.

Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genetic Epidemiology. 2000; 19:S36–S42. [PubMed: 11055368]

Lange C, Laird NM. Power calculations for a general class of family-based association tests: Dichotomous traits. American Journal of Human Genetics. 2002; 71:575–584. [PubMed: 12181775]

Lange C, DeMeo DL, Laird NM. Power and design considerations for a general class of family-based association tests: Quantitative traits. American Journal of Human Genetics. 2002; 71:1330–1341. [PubMed: 12454799]

Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, Becker J, Herberz R, AlChawa T, Nasser E, Boehmer AC, Mattheisen M, Alblas MA, Barth S, Kluck N, Lauster C, Braumann B, Reich RH, Hemprich A, Poetzsch S, Blaumeiser B, Daratsianos N, Kreusch T, Murray JC, Marazita ML, Ruczinski I, Scott AF, Beaty TH, Kramer FJ, Wienker TF, Steegers-Theunissen RP, Rubini M,

Mossey PA, Hoffmann P, Lange C, Cichon S, Propping P, Knapp M, Noethen MM. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nature Genetics. 2012; 44:968–971. [PubMed: 22863734]

Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Human Heredity. 2000; 50:211–223. [PubMed: 10782012]

Rao, CR. Linear Statistical Inference and its Applications. 2 edition. New York: Wiley & Sons; 1973.

Schaid DJ. General score tests for associations of genetic markers with disease using cases and their parents. Genetic Epidemiology. 1996; 13:423–449. [PubMed: 8905391]

Schaid DJ. Likelihoods and TDT for the case-parents design. Genetic Epidemiology. 1999; 16:250–260. [PubMed: 10096688]

Schaid DJ, Sommer SS. Comparison of statistics for candidate-gene association studies using cases and parents. American Journal of Human Genetics. 1994; 55:402–409. [PubMed: 8037216]

Schwender H, Taub MA, Beaty TH, Marazita ML, Ruczinski I. Rapid testing of SNPs and gene-environment interactions in case-parent trio data based on exact analytic parameter estimation. Biometrics. 2012; 68:766–773. [PubMed: 22150644]

Self SG, Longton G, Kopecky KJ, Liang KY. On estimating HLA/disease association with application to a study of aplastic anemia. Biometrics. 1991; 47:53–61. [PubMed: 2049513]

Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. American Journal of Human Genetics. 1996; 59:983–989. [PubMed: 8900224]

Spielman RS, McGinnis R, Ewens WJ. Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). American Journal of Human Genetics. 1993; 52:506–516. [PubMed: 8447318]

Table 1

Number of minor alleles shown by the affected offspring, their parents, and the corresponding pseudo-controls for the ten possible genotype combinations case-parent trios can show as well as the weights of the trios in the maximimization of the conditional likelihood assuming an additive mode of inheritance. The number of trios showing a specific genotype combination is denoted by $n_c^{(p_1,p_2)}$ with $c, p_1, p_2 \in \{0, 1, 2\}$ and $p_1 \leq p_2$. This is a modified version of Table 1 from Schwender et al. (2012).

Number of Trios	Number of Minor Alleles			Weights in the Likelihood
	Parents	Offspring	Pseudo-Controls	
$n_0^{(0,1)}$	0, 1	0	0, 1, 1	$\dfrac{1}{2+2\exp(\gamma_{\mathrm{add}})}$
$n_1^{(0,1)}$	0, 1	1	0, 0, 1	$\dfrac{\exp(\gamma_{\mathrm{add}})}{2+2\exp(\gamma_{\mathrm{add}})}$
$n_1^{(1,2)}$	1, 2	1	1, 2, 2	$\dfrac{1}{2+2\exp(\gamma_{\mathrm{add}})}$
$n_2^{(1,2)}$	1, 2	2	1, 1, 2	$\dfrac{\exp(\gamma_{\mathrm{add}})}{2+2\exp(\gamma_{\mathrm{add}})}$
$n_0^{(1,1)}$	1, 1	0	1, 1, 2	$\dfrac{1}{(1+\exp(\gamma_{\mathrm{add}}))^2}$
$n_1^{(1,1)}$	1, 1	1	0, 1, 2	$\dfrac{\exp(\gamma_{\mathrm{add}})}{(1+\exp(\gamma_{\mathrm{add}}))^2}$
$n_2^{(1,1)}$	1, 1	2	0, 1, 1	$\dfrac{\exp(2\gamma_{\mathrm{add}})}{(1+\exp(\gamma_{\mathrm{add}}))^2}$
$n_1^{(0,2)}$	0, 2	1	1, 1, 1	$\dfrac{1}{4}$
$n_2^{(2,2)}$	2, 2	2	2, 2, 2	$\dfrac{1}{4}$
$n_0^{(0,0)}$	0, 0	0	0, 0, 0	$\dfrac{1}{4}$

Table 2

Sample sizes required to gain 80% power with an type I error rate of $\alpha = 5 \times 10^{-8}$ for genotypic TDTs and score tests assuming different modes of inheritance.

RR	MAF	Additive		Dominant		Recessive	
		gTDT	Score	gTDT	Score	gTDT	Score
1.05	0.01	1,642,576	1,642,381	1,662,620	1,662,423	216,037,547	216,012,183
	0.10	183,105	183,084	207,695	207,672	2,310,954	2,310,703
	0.20	104,520	104,508	136,124	136,110	631,272	631,207
	0.50	69,857	69,849	154,609	154,594	149,732	149,718
1.20	0.01	110,732	110,549	111,955	111,770	14,148,787	14,125,709
	0.10	12,821	12,800	14,360	14,338	152,006	151,777
	0.20	7,621	7,610	9,657	9,643	41,833	41,774
	0.50	5,704	5,696	11,620	11,604	10,310	10,296
1.30	0.01	51,639	51,463	52,174	51,996	6,482,074	6,460,251
	0.10	6,123	6,103	6,807	6,785	69,827	69,610
	0.20	3,730	3,718	4,651	4,637	19,308	19,252
	0.50	2,976	2,968	5,790	5,773	4,876	4,863
1.40	0.01	30,425	30,255	30,722	30,549	3,755,909	3,735,177
	0.10	3,690	3,671	4,075	4,054	40,564	40,358
	0.20	2,300	2,289	2,827	2,813	11,268	11,215
	0.50	1,942	1,934	3,629	3,612	2,913	2,900
1.50	0.01	20,363	20,198	20,550	20,383	2,474,435	2,454,659
	0.10	2,524	2,505	2,771	2,749	26,789	26,592
	0.20	1,607	1,596	1,950	1,935	7,475	7,424
	0.50	1,427	1,419	2,575	2,557	1,977	1,964

Table 3

Sample sizes required to gain 80% power with an type I error rate of $\alpha = 10^{-7}$ for the genotypic TDTs, the score tests, and the first approach to power approximation proposed by Knapp (1999) assuming a dominant or a recessive mode of inheritance. The sample sizes for the method by Knapp (1999) were obtained from Table 3 in Knapp (1999).

RR	MAF	Dominant			Recessive		
		gTDT	Score	Knapp (1999)	gTDT	Score	Knapp (1999)
1.5	0.01	19,741	19,582	19,755	2,378,326	2,359,578	154,174,890
	0.10	2,661	2,641	2,897	25,747	25,561	174,694
	0.50	2,473	2,456	4,568	1,900	1,887	3,099
	0.80	11,636	11,553	50,826	2,047	2,032	2,356
2.0	0.01	6,031	5,890	5,947	680,674	665,193	38,654,522
	0.10	877	857	949	7,449	7,294	45,071
	0.50	969	950	1,839	622	610	957
	0.80	4,819	4,717	21,998	762	746	851
4.0	0.01	1,215	1,102	1,115	115,532	105,324	4,344,070
	0.10	225	204	231	1,306	1,200	5,631
	0.50	361	331	696	158	146	207
	0.80	1,975	1,802	9,384	256	234	259

Table 4

Simulated power of the genotypic TDT and the score test. The power was determined assuming either an additive or a dominant mode of inheritance and based on 10^5 simulated data sets consisting of the number of trios summarized in Table 2 for each of the settings in Table 2.

RR	MAF	Genotypic TDT		Score Test	
		Additive	Dominant	Additive	Dominant
1.05	0.01	0.801	0.801	0.800	0.800
	0.10	0.799	0.798	0.800	0.802
	0.20	0.801	0.800	0.799	0.799
	0.50	0.801	0.800	0.800	0.799
1.20	0.01	0.800	0.800	0.801	0.801
	0.10	0.801	0.799	0.799	0.800
	0.20	0.799	0.798	0.798	0.801
	0.50	0.799	0.801	0.799	0.800
1.30	0.01	0.800	0.800	0.800	0.800
	0.10	0.798	0.802	0.799	0.798
	0.20	0.799	0.800	0.801	0.798
	0.50	0.800	0.800	0.798	0.800
1.40	0.01	0.800	0.798	0.800	0.801
	0.10	0.800	0.798	0.800	0.800
	0.20	0.799	0.799	0.800	0.800
	0.50	0.802	0.800	0.799	0.801
1.50	0.01	0.800	0.803	0.798	0.799
	0.10	0.798	0.799	0.800	0.801
	0.20	0.797	0.800	0.802	0.798
	0.50	0.798	0.800	0.798	0.798

Table 5

Number of minor alleles shown by the affected offspring and their parents for the ten possible genotype combinations case-parent trios can exhibit as well as the codings for the variable in the conditional logistic regression model used to perform a genotypic TDT assuming either a dominant or recessive mode of inheritance. The number of trios showing a specific genotype combination is denoted by $n_c^{(p_1,p_2)}$ with c, p_1, $p_2 \in \{0, 1, 2\}$ and $p_1 \leq p_2$. Additionally, the corresponding components of the genotypic TDT statistic used to derive the asymptotic normal distribution of this statistic are displayed, where – marks a genotype combination that does not contribute to the maximization of the conditional log-likelihood under the corresponding mode of inheritance.

Number of Trios	Number of Minor Alleles		Coding for Case/Pseudo-Controls		Components	
	Parents	Offspring	Dominant	Recessive	Dominant	Recessive
$n_0^{(0,1)}$	0, 1	0	0 / 0, 1, 1	0 / 0, 0, 0	d_1	–
$n_1^{(0,1)}$	0, 1	1	1 / 0, 0, 1	0 / 0, 0, 0	d_2	–
$n_1^{(1,2)}$	1, 2	1	1 / 1, 1, 1	0 / 0, 1, 1	–	r_1
$n_2^{(1,2)}$	1, 2	2	1 / 1, 1, 1	1 / 0, 0, 1	–	r_2
$n_0^{(1,1)}$	1, 1	0	0 / 1, 1, 1	0 / 0, 0, 1	d_3	r_3
$n_1^{(1,1)}$	1, 1	1	1 / 0, 1, 1	0 / 0, 0, 1	d_4	r_3
$n_2^{(1,1)}$	1, 1	2	1 / 0, 1, 1	1 / 0, 0, 0	d_4	r_4
$n_1^{(0,2)}$	0, 2	1	1 / 1, 1, 1	0 / 0, 0, 0	–	–
$n_2^{(2,2)}$	2, 2	2	1 / 1, 1, 1	1 / 1, 1, 1	–	–
$n_0^{(0,0)}$	0, 0	0	0 / 0, 0, 0	0 / 0, 0, 0	–	–