RESEARCH NOTE

Sample size computation for association studies using case–parents design

NAJLA KHARRAT, IMEN AYADI and AHMED REBAÏ*

Unité de Biostatistique et de Bioinformatique, Centre de Biotechnologie de Sfax, B.P. 'K', 3038 Sfax, Tunisia

Introduction

One of the most common designs for identification of genes involved in complex human diseases is the transmission disequilibrium test (TDT) design (Spielmann *et al*. 1993). This design consists in the collection of cases (affected individuals) and their two parents (commonly called triads of case–parents). The analysis of such data is based on comparison of the numbers of transmissions and nontransmissions of alleles for a candidate locus from parents to affected child. Here we mean by a candidate locus any variant or marker near or within the coding sequence of a candidate gene, whose function or position (or both) suggest its involvement in disease aetiology.

Differential transmission of alleles, if any, indicates linkage and association between the candidate locus and the disease. Consider, without loss of generality, a candidate locus with two alleles A and a (in the case of multiple alleles such as for STR markers, one would consider testing each allele A against all others pooled in one non-A (a) allele). The expression of the TDT statistic is $T = (b - c)^2/(b + c)$, where b (respectively c) is the number of heterozygous parents who transmit the A (respectively a) allele to their affected child. Thus, for a triad to be informative (contribute to T value), at least one parent needs to be heterozygous for the candidate locus. TDT is a test of both linkage and association; the null hypothesis is linkage and/or association and it was shown that the expectation of TDT under the alternative hypothesis is proportional to $(1-2\theta)\delta$, where θ is the recombination fraction between the candidate locus and the disease locus and δ the linkage disequilibrium between the two loci.

Under the null hypothesis, heterozygous parents transmit either allele (A, a) to an affected offspring with equal probability. For a large number of triads (>20), the distribution of T can be approximated by a chi-square distribution with one degree of freedom (d.f.). Many approximations have been proposed to compute power of the TDT statistic for a given sample size and gene parameters or conversely to get sample size needed to reach a given power (Knapp 1999; Schaid 1999; Chen and Deng 2001; Brown 2004). In their seminal paper, Risch and Merikangas (1996) showed that for a multiplicative mode of inheritance (MOI) for the susceptibility gene, sample size depends on two parameters: the frequency of the risk allele at the candidate locus (p) and the genotypic relative risk (GRR) attributable to that locus. GRR could be expressed as a function of a single parameter, denoted γ, depending on the MOI (Schaid 1999). Various analytical approaches based on the normal distribution of the square root of the TDT statistic have since been proposed and have yielded comparable results (Chen and Deng 2001; Iles 2002). However, none of these methods has provided a simple expression of sample size as a function of p and γ.

Deng and Chen (2001) showed that under the alternative hypothesis the test statistic TDT follows approximately a noncentral chi-square. They gave the general expression of the noncentrality parameter as a function of sample size, allele frequencies, penetrances and prevalence of the candidate locus.

Here we build on their work and derive analytical expressions of the noncentrality parameter as a function of p and γ for different MOI. Using approximations for the noncentral chi-square quantile computation we provide expressions of sample size N as simple functions of p and γ, for a given power and significance level and for different MOI. The performance of our approximations is assessed by comparison to other methods. Extension of our approximations to compute sample size for case–control association studies is discussed.

Materials and methods

Parameter settings

We consider a candidate locus with two alleles A and a where A is putatively associated with the disease status (increasing

*For correspondence. E-mail: Ahmed.Rebai@cbs.rnrt.tn.

Keywords. sample size; association tests; genotype relative risk; power; autism.

disease risk). The frequency of A in the general population is p. We denote by f_0, f_1 and f_2 the penetrance or prevalence (the probability of developing the disease when having a genotype) of genotypes *aa*, *Aa* and *AA*, respectively. The genotypic risk ratio (GRR) of a given genotype is defined as the ratio of its prevalence to that of a reference genotype (with no copy of the risk allele, here *aa*). We thus have to specify two GRRs ($g_0 = 1$) : $g_1 = f_1/f_0$ and $g_2 = f_2/f_0$ that can be expressed according to a single parameter (denoted γ) depending on the MOI (table 1). For complex diseases, where MOI is generally unknown, four common models are generally considered in power studies of TDT: recessive, dominant, additive and multiplicative. Therefore, only two parameters need to be specified in power computations: γ and p. Note that $\gamma = 1$ should correspond to no effect of the candidate locus and all GRRs should be equal in this case. This explains why we used $2\gamma - 1$ for the additive model (as in Schaid 1999) and not 2γ as used by other authors (Camp 1999; Knapp 1999).

Analytical approximation

Deng and Chen (2001) have shown that the TDT statistic follows approximately a noncentral chi-square distribution under the alternative hypothesis of linkage and association. Using similar computation to those of Nielsen *et al*. (1998) we can get the general expression of the noncentrality parameter λ as a function of sample size N, frequency p, and genotypes and disease prevalence as below:

$$\lambda = N\frac{2p(1-p)[pg_2 + (1-2p)g_1 - (1-p)]^2}{[pg_2 + g_1 + (1-p)][p^2g_2 + 2p(1-p)g_1 + (1-p)^2]}. \quad (1)$$

Starting from this expression and using table 1, we found, after some algebraic developments using Maple®, the following expressions of λ for the multiplicative and additive MOI (similar expressions are obtained for other MOI):

$$\text{Multiplicative: } \lambda_{\mathrm{M}} = N\frac{2p(1-p)(\gamma-1)^2}{[1+p(\gamma-1)](\gamma+1)} \quad (2)$$

$$\text{Additive: } \lambda_{\mathrm{A}} = N\frac{2p(1-p)(\gamma-1)^2}{[1+\gamma+2p(\gamma-1)][2p(\gamma-1)+1]}. \quad (3)$$

Using the approximation proposed by Winer *et al*. (1991) for a noncentral chi-square with one d.f. and noncentrality parameter λ, the power at a significance level α can be computed as

$$\pi = \Pr(\chi^2_{(1,\lambda)} > C_\alpha) \approx \Pr(Z > Z_0),$$

where C_a is the α-quantile of a central chi-square with 1 d.f., Z a standard normal deviate and Z_0 the π-quantile of the normal distribution. Z_0 and λ are related by

$$Z_0 = \frac{\sqrt{2C_a(1+\lambda)} - \sqrt{\lambda + (\lambda+1)^2}}{\sqrt{1+2\lambda}}. \quad (4)$$

For a power of $\pi = 80\%$ we have $Z_0 = -0.84$. We can obtain λ for any given power and significance level by numerically solving equation 4. Table 2 gives values of C_α and λ for different α values and for 80% power, calculated using Maple®.

From equations (2) and (3) we computed N for different values of p, ranging from 0.01 (rare) to 0.5 (common), and γ (ranging from 1.5 to 4). We considered a power of 80% and $\alpha = 5 \times 10^{-8}$ to allow comparison with other published results (see Discussion).

Performance assessment

To assess the performance of the approximations provided by the analytical equations above, we computed the number of triads needed, N, for the same sets of parameters using the numerical method implemented in TDT Power Calculator (TDT-PC) computer program (Chen and Deng 2001). We also calculated simulated power (10,000 replications) corresponding to sample sizes obtained with our formulas for many parameter combinations (p = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, and γ = 1.5, 2.0, 2.5, 3.0, 3.5, 4.0). This power should be close to 80% for our approximation to be a good method for sample size computation.

Table 1. Genetic models and their respective genotypic relative risks.

	Mode of inheritance			
Genotype relative risk	Multiplicative	Additive	Recessive	Dominant
g_1 (for *Aa* genotype)*	γ	γ	1	γ
g_2 (for *AA* genotype)	γ^2	$2\gamma - 1$	γ	γ

*A is the risk allele.

Table 2. Value of noncentrality parameter λ for different values of significance level α and for 80% power.

α	0.05	0.01	0.005	0.001	5×10^{-4}	10^{-4}	5×10^{-5}	10^{-6}	5×10^{-6}	5×10^{-8}
C_α*	3.841	6.634	7.879	0.827	12.115	15.136	16.448	19.511	20.837	29.716
λ	13.584	21.248	24.519	32.046	35.263	42.687	45.867	53.211	56.358	77.104

*α-Quantile of a central chi-square distribution with 1 degree of freedom.

Results

Sample size calculations and comparison with other methods

Studying mathematical properties of N as a function of p and γ we found that $N(p, \gamma)$ is an increasing function of both parameters (for $\gamma > 1$). Evolution of sample size N for the multiplicative and additive MOI are reported in figures 1 and 2. Comparing our results for different values of γ and p with results obtained from the method of Chen and Deng (2001) based on numerical integration, we found that sample sizes are very close to each other (within 3%) on either side, indicating absence of bias in the comparison (results not shown). Similar results were found for all four MOI (in table 1).

Figure 1 gives the evolution of N according to γ for $p = 0.1$ and $p = 0.5$ and for multiplicative and additive MOI. Figure 2 depicts evolution of N as a function of p for $\gamma = 2$. We see a typical U-shaped function. In fact, using analytical derivations one can show that N, for any fixed γ, has a minimum at frequency

$$p = (1 + \sqrt{\gamma})^{-1} \quad \text{for multiplicative MOI,} \tag{5}$$

$$p = \sqrt{6\gamma^3 + \gamma^2 - 4\gamma + 1} \Big/ 6\gamma(\gamma - 1) \quad \text{for additive MOI.} \tag{6}$$

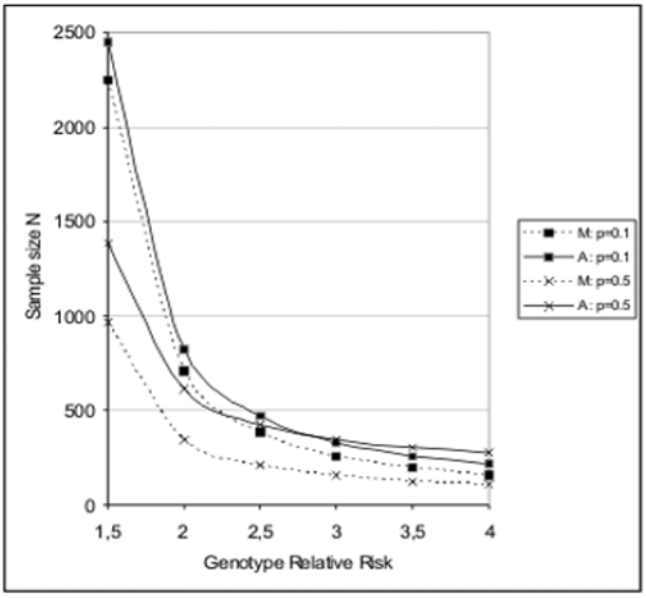

Figure 1. Sample size as a function of GRR parameter γ and for $p = 0.1$ and $p = 0.5$ and for both multiplicative (M) and additive (A) MOI.

Values of these minima vary from 0.45 to 0.33 and from 0.37 to 0.20 when γ varies from 1.5 to 4, respectively. They are 0.41 and 0.31 for $\gamma = 2$ (figure 2).

From figures 1 and 2, one can see that sample sizes obtained under the multiplicative MOI are always smaller than those found under the additive MOI. The difference increases when γ decreases (for $\gamma > 1$) and p increases.

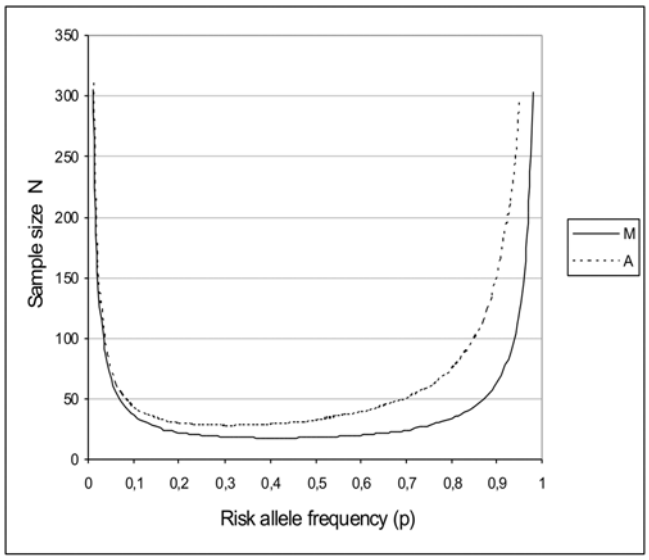

Figure 2. Sample size as a function of frequency p for $\gamma = 2$ for additive (A) and multiplicative (M) MOI.

As consistently reported by all methods, we see from our data that for rare alleles ($p < 0.10$) and small GRR ($\gamma < 2$) more than 2200 triads are needed to achieve 80% power at 5×10^{-8} level; TDT, like most commonly used association tests, is inefficient. For non-rare and common alleles ($0.1 \leq p \leq 0.50$), often reported as risk alleles for complex diseases, sample sizes needed are in the range 600–2000 for low GRR ($\gamma \leq 2$) and 100–350 for relatively high GRR ($\gamma > 2$).

Precision

The average simulated power for the parameter combinations studied was 80.9% and 79.6% for multiplicative and additive MOI, respectively. The corresponding standard deviations are 0.0093 and 0.0081, showing that simulated power is very close to the expected value of 80%. Note that it is only for high GRR ($\gamma = 4$) and the multiplicative model that the simulated power is always larger than 80% (84.3% ± 3.1%). Thus for low and medium GRR ($1.5 \leq \gamma \leq 3.50$) our method gives correct power.

Discussion

Comparison with other methods

Among all previously described sample size computation methods for TDT, our method has the great advantage of simplicity, while keeping good performance compared to others. Our equations (2) and (3) could be easily implemented in any programming language. For commonly used significance levels, the values of parameter λ could be obtained from our table 2.

Another expression of N has been proposed by Schaid (1996) for multiplicative MOI, based on a formulation of

TDT as a general score test: $\lambda_M = N\frac{\log(\gamma)^2 2p(1-p)\gamma}{[1+p(\gamma-1)](\gamma+1)}$. This expression is derived for log-additive model and small GRR. As expected, it reduces (at the first-order Taylor series expansion since $\log(\gamma) \approx \gamma - 1$ for γ close to 1) to our expression (2) for small GRR and gives very close sample sizes for $\gamma \leq 2$.

Practical use of our approximation for sample size computation

Most previous studies that addressed sample size or power computation have focussed on evaluation and comparison of performance of the methods rather than on practical recommendations for experimenters on how to use them in designing TDT studies. Even available programs are not easy to use by experimenters. A particular problem is that before any sample size calculation one should choose which MOI to consider, and should give a value for allele frequency and GRRs, whereas in practice all these parameters are unknown. If one considers that the disease locus in not the marker itself (incomplete linkage or linkage disequilibrium) the computations become even more problematic. Below we give some practical guidelines for sample size computation using our method.

How to choose risk allele frequency: The HapMap project (The International HapMap Consortium 2003) provides data on thousands of single-nucleotide polymorphism (SNP) markers throughout the genome. One can find which are the best SNP markers to use (depending on the origin of the population studied: African, European, Asian) and their allele and haplotype frequencies. So if one is interested in testing a candidate locus for association with a common disease, one could use an average allele frequency of the minor alleles of the considered SNPs or even the frequency of the haplotypes of this allele. Most alleles reported to be associated with increasing risk in complex diseases have frequencies in the range 0.1–0.5. So, if experimenters have no idea of the risk allele frequency, they can either use $p = 0.3$ as an average or compute the minimal sample size needed for any given γ (the corresponding frequencies are thus obtained from our expressions (5) and (6) and are in the range 0.20–0.50).

How to choose mode of inheritance (MOI): Risch and Merikangas (1996) advocated the use of the multiplicative MOI for calculating sample sizes. This model is, in fact, a convenient approximation for complex diseases because the tests based on the multiplicative model (even if the true MOI is not) remain valid in the sense of having correct size (because of independence of parental transmissions). However, it may not be optimal in all situations in the sense of being the most powerful and the sample sizes it provides are underestimates (Schaid 1999). Moreover, results from many association studies show that the additive MOI is better suited to describe the effect of susceptibility genes of complex diseases (the multiplicative model attributes extreme GRR to the *AA* genotype). We propose here that computations be done for both models providing a sample size interval within which the experimenter may choose.

Which power and significance level to use: There is a consensus on the use of 80% power for sample size but less stringent power (for example 50%, which corresponds to missing or finding a true positive with equal chances) could be used if one is looking for association with a candidate locus for the first time. A significance level of 5×10^{-8} is used in most studies. This value was proposed in Risch and Merikangas (1996) on the basis of a Bonferroni correction accounting for five diallelic markers being tested in each of 100,000 genes (one million tests). However, since the total number of genes in the human genome is currently estimated in the range 25,000–35,000 (Pennisi 2003) and recent studies (Goldstein *et al*. 2003) show that 150,000 representative SNPs are enough for a genome scan in European populations, a level of about 2×10^{-7} may be more appropriate. If we consider that a genewide test is performed to ensure a genomewide level of 5% for a genome with 30,000 genes (all of which are potentially candidates) and a disease having 10 risk genes, a per test level of 5×10^{-5} is more appropriate (van den Oord and Sullivan 2003). We recommend this level for future sample size calculations.

For which GRR: Most candidate genes that were reported to be associated with complex diseases have low to moderate GRR ($1.5 \leq \gamma \leq 3$). A value of $\gamma = 2$ is, in this context, a good compromise. However, if sample size computations are to be done for a replication study, after an initial positive or negative report, one can roughly estimate γ by the ratio b/c, where b and c are transmission and nontransmission respectively (Schaid 1999).

Generalization to case–control sample size computation

Long *et al*. (1997) gave the expression, under multiplicative MOI, of the noncentrality parameters of the chi-square test used in case–control association studies. We derived similar expressions for other MOI and used our approach (based on approximation (4)) to obtain the expression of sample size as a function of p, γ and disease prevalence.

Implementation

A Microsoft Excel worksheet implements our approach and computes sample sizes for 80% power, for a given γ or p or both and a chosen significance level α. This worksheet also implements our method for case–control designs and is available on request. Another implementation, in the R language (http://www.r-project.com/) as part of an R package containing tools for power and sample size computation for many genetic problems, is also available.

Application

We planned to study association of the *DSS1* gene (MIM 601285) in autistic disorder (MIM 209850) with a TDT design. A first investigation with 28 triads and marker D7S479 showed significant association, although the number of informative transmissions was very small (only 5). We explored the HapMap data and found eight SNP within the *DSS1* coding sequence. Frequencies of the minor allele are in the range 0.12–0.43 with an average of 0.28. We thus computed sample size with $p = 0.28$, $\gamma = 2$, 80% power and $\alpha = 5 \times 10^{-5}$. We found that we need $N = 218$ and 315 trios for multiplicative and additive MOI, respectively. For $\alpha = 5\%$ the sample sizes were 65 and 93, respectively.

We think that the sample sizes obtained by our approach and according to the guidelines we provide would be of great help to human geneticists in determining sample size requirements when designing their association studies.

Acknowledgements

This work was supported by the Ministry of Scientific Research, Technology and Skills Development, Tunisia. We would like to thank Dr H. Hadj-Kacem whose questions and discussions about sample size computation for TDT studies have motivated this work.

References

Brown B. W. 2004 Power calculations for the transmission/disequilibrium and affected sib pair tests using elementary probability methods. *Genet. Res.* **83**, 133–141.

Camp N. J. 1999 Genomewide transmission/disequilibrium testing: a correction. *Am. J. Hum. Genet.* **64**, 1485–1487.

Chen W. M. and Deng H. W. 2001 A general approach for computing the statistical power of transmission/disequilibrum test for complex disease genes. *Genet. Epidemiol.* **21**, 53–67.

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

Goldstein D. B., Ahmadi K. R., Weale M. E. and Wood N. W. 2003 Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet.* **19**, 615–622.

Iles M. M. 2002 On calculating the power of a TDT study—comparison of methods. *Ann. Hum. Genet.* **66**, 323–328.

The International HapMap Consortium 2003 The International HapMap Project. *Nature* **426**, 789–796.

Knapp M. A. 1999 Note on power approximations for the transmission/disequilibrium test. *Am. J. Hum. Genet.* **64**, 1177–1185.

Long A. D., Grote M. N. and Langley C. H. 1997 Genetic analysis of complex diseases. *Science* **275**, 1328.

Nielsen D. M., Ehm M. G. and Weir B. S. 1998 Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.* **63**, 1531–1540.

Pennisi E. 2003 Gene counters struggle to get the right answer. *Science* **301**, 1040–1041.

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

Schaid D. J. 1996 General score tests for associations of genetic markers with disease using cases and their parents. *Genet. Epidemiol.* **13**, 423–449.

Schaid D. J. 1999 Likelihoods and TDT for the case-parents design. *Genet. Epidemiol.* **16**, 261–273.

Spielmann R. C., McGinnis R. E. and Ewens W. J. 1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516.

van den Oord E. J. and Sullivan P. F. 2003 A framework for controlling false discovery rates and minimizing the amount of genotyping in the search for disease mutations. *Hum. Hered.* **56**, 188–199.

Winer B. J., Brown D. R. and Michels K. M. 1991 *Statistical principles in experimental design*, 3rd edition. McGraw-Hill, New York.

Received 28 April 2005