J Hum Genet (2005) 50:46–48
DOI 10.1007/s10038-004-0217-5

SHORT COMMUNICATION

Karen L. Ayres · Robert N. Curnow

Detecting non-multiplicative genotype relative risks from transmissions of parental alleles to affected children

Received: 23 July 2004 / Accepted: 25 October 2004 / Published online: 14 December 2004

Abstract The differential transmission of alleles from parents to affected children indicates that the locus under investigation is either directly involved in the occurrence of the disease or that there are allelic associations with other loci that are directly involved. Conditional logistic regression applied to a diallelic locus leads to a test with two degrees of freedom. The power of a single degree of freedom test to detect non-multiplicative allelic effects is discussed here.

Keywords Genotype relative risks · Conditional logistic regression · Allelic transmission

The differential transmission rates of alleles from heterozygous parents to children affected by a disease can provide information about the likely location and mode of action of genes affecting susceptibility to that disease. Conditional logistic regression (CLR), first proposed by Self et al. (1991), is based on a comparison of the frequency of pairs of alleles inherited by affected children with the other combinations of the alleles that they might have inherited. Here we consider the power associated with fitting the second term in the usual CLR model after a first multiplicative term has already been fitted. Testing for a non-multiplicative effect of the alleles is needed to justify the use of a single multiplicative parameter to summarise the relationship of the alleles to the occurrence of the disease. Also, increasing interest in the biochemical pathways through which the effects of the genotypes are mediated requires knowledge of the quantitative relationship between specific genotypes and the phenotypic expression of the disease. A non-multiplicative effect at a marker locus would indicate non-multiplicative effects at an associated disease locus although, with no knowledge of the level of the allelic associations, their size would be unknown.

For generality, we shall refer to the locus as a marker locus so that a candidate gene is a special case. We denote the alleles at a diallelic marker by M_1 and M_2, and their frequencies by m and $(1-m)$, respectively. The CLR approach involves treating the affected children as "cases" and the three genotypes formed by the un-inherited alleles as matched "controls". If the log relative risks are assumed to be linear in variates x_1 and x_2 that represent the child's marker genotype, then the probability of the disease p_i for child i is assumed to be related to x_1 and x_2 by the logistic regression equation (e.g. Collett 2003)

$$\log \frac{p_i}{1-p_i} = \alpha_i + \beta_1 x_{1i} + \beta_2 x_{2i}$$

The genotype relative risks of the marker genotypes M_1M_1, M_1M_2 and M_2M_2 are denoted r_2, r_1 and 1. If the marker locus is not linked to any locus affecting susceptibility to the disease or if there is no association between the alleles at the marker locus and the alleles at the disease loci then $r_2 = r_1 = 1$. The test of no differential transmission, and therefore no association between the locus and the disease, has two degrees of freedom. Many authors have studied the power of this test and the power of the single degree of freedom tests based on specific single parameter models representing dominant ($r_2 = r_1$), recessive ($r_1 = 1$) and multiplicative ($r_2 = r_1^2$) genotype relative risks (e.g. Schaid and Sommer 1993, 1994; Spielman et al. 1993; Schaid 1996; Sham 1998; Schaid 1999). The multiplicative model is thought to be the best single parameter model in terms of representing alternative models such as additive, dominant and recessive (Schaid 1996). We present results on the power of the one degree of freedom test of the non-multiplicative term in the regression, conditional on having fitted a multiplicative term.

This paper is dedicated to the memory of Professor Steve Bennett of the London School of Hygiene and Tropical Medicine who was to be involved in the work on which it is based.

K. L. Ayres · R. N. Curnow (✉)
School of Applied Statistics, The University of Reading,
P.O. Box 240, Earley Gate, Reading, RG6 6FN, UK
E-mail: r.n.curnow@reading.ac.uk
Tel.: +44-118-3788022
Fax: +44-118-9753169

Following Schaid (1999), we simulated families with affected children using the null, additive, multiplicative, dominant and recessive models with the genotype relative risks $r_1 = 2$ and 4, with frequencies $m = 0.1$ and 0.5. Marker alleles positively associated with disease alleles are unlikely to have frequencies higher than 0.5. Random mating was assumed in generating the parental mating types. The total number of families was set at $n = 100$ or 200. Only families with at least one parent heterozygous are informative. The probability P that a family with an affected child is informative is given by

$$P = \frac{m^2(1-m^2)r_2 + 2m(1-m)(1-m(1-m))r_1 + (1-m)^2(1-(1-m)^2)}{m^2 r_2 + 2m(1-m)r_1 + (1-m)^2}$$

The expected number of informative families in a study of size n is therefore nP.

For the multiplicative model, x_1 took the values 2, 1 and 0 for the genotypes M_1M_1, M_1M_2 and M_2M_2, respectively. Since a model with x_1 and x_2 is a full model, the variable x_2 can take any non-additive values; we used 1, 1 and 0 for the above genotypes, respectively.

For each of 10,000 simulations, we fitted the CLR models using the function clogit in the survival package of the statistical program R (Ihaka and Gentleman 1996). We first tested the full model with x_1 and x_2 and the reduced model with x_1 only and recorded the number of results significant at 5% using a likelihood ratio test. These results were as expected, given the work of other authors (e.g. Schaid 1996, 1999) and are omitted. However, we also performed a test based on the difference of deviances for these two models, as a test of deviations from the multiplicative models, and these results are presented here.

The regression analysis does not converge when the data do not allow the separation of the effects of the different parameters or deviate strongly from the pattern predicted by the model. Table 1 presents, for all genetic models considered, the proportion of the analyses of the full model that converged, together with the attributable risk and the expected proportion of families with an affected child that are informative, P. The attributable risk is defined as the population lifetime prevalence of the disease minus the disease penetrance for the least disease-related genotype, M_2M_2, as a proportion of the population lifetime prevalence. In our notation, the attributable risk is $1-1/(m^2 r_2 + 2m(1-m)\, r_1 + (1-m)^2)$.

When $m = 0.5$, 75% of families are expected to be informative, but this figure is much lower for $m = 0.1$. The convergence rate for the analysis of the full model with $n = 100$ is at least 0.60 when $m = 0.1$ and 0.95 when $m = 0.5$. With larger samples, convergence is almost certain for most models with both $m = 0.1$ and 0.5. Convergence for the model involving x_1 only is almost identical to that for the full model.

Table 1 also shows the power, calculated from the simulations that converged, of the test at a 5% significance level based on the additional variation explained by fitting x_2 having already fitted x_1, a test indicated in Table 1 by $x_2|x_1$. This test has the correct power, 0.05, when the null or the multiplicative model holds and has power of at least 30% when $m = 0.1$ for all the stronger models except the additive model when $n = 100$. The power is very low for all the weaker models. When $m = 0.5$, the power increases to about 30% for the stronger dominance and recessive models even with $n = 100$, but the power for the weaker additive model is again very low.

As expected, the predictions of the weak additive model are similar to those of the multiplicative model. Otherwise, there is sufficient power in the test to suggest that the regression on the multiplicative term x_1 only should always be calculated and the amount of variation explained by the full model then compared with the amount explained by this single term regression.

Table 1 Properties of the nine genetic models considered, including attributable risk values (Schaid and Sommer 1993), expected proportion of informative families in a sample (P) and the percentage of simulations (out of 10,000) that reached convergence for the logistic regression model involving parameters x_1 and x_2, for $n = 100$ and $n = 200$ simulated families. *M* multiplicative, *A* additive, *D* dominance, *R* recessive

Model	Genotype relative risks		Attributable risk		P		Percentage of simulations reaching convergence ($n = 100$, $n = 200$)		Power of test[a] $x_2\|x_1$ ($n = 100$, $n = 200$)	
	r_2	r_1	$m = 0.1$	$m = 0.5$	$m = 0.1$	$m = 0.5$	$m = 0.1$	$m = 0.5$	$m = 0.1$	$m = 0.5$
Null	1	1	0	0	0.33	0.75	0.60, 0.86	1.00, 1.00	0.052, 0.069	0.050, 0.051
2M	4	2	0.17	0.56	0.43	0.75	0.95, 1.00	1.00, 1.00	0.069, 0.054	0.056, 0.053
2A	3	2	0.17	0.50	0.43	0.75	0.91, 0.99	1.00, 1.00	0.089, 0.081	0.099, 0.138
2D	2	2	0.16	0.43	0.42	0.75	0.80, 0.96	1.00, 1.00	0.140, 0.181	0.311, 0.540
2R	2	1	0.01	0.20	0.33	0.75	0.83, 0.98	1.00, 1.00	0.107, 0.155	0.296, 0.528
4M	16	4	0.41	0.84	0.57	0.75	1.00, 1.00	0.95, 1.00	0.056, 0.056	0.068, 0.051
4A	7	4	0.38	0.75	0.55	0.75	0.99, 1.00	0.99, 1.00	0.201, 0.340	0.324, 0.546
4D	4	4	0.36	0.69	0.54	0.75	0.91, 0.99	1.00, 1.00	0.398, 0.652	0.784, 0.971
4R	4	1	0.03	0.43	0.35	0.75	0.95, 1.00	1.00, 1.00	0.323, 0.556	0.710, 0.944

[a]Powers are for the test based on fitting x_2 given that x_1 has already been fitted at a 5% level of significance (based on the convergent simulations)

Depending on the results of the test for non-multiplicative effects, the data can then be summarised in terms of estimates, together with confidence intervals, of either the relative risks for all three genotypes or the multiplicative effect of the allele.

References

Collett D (2003) Modelling binary data, 2nd edn. Chapman and Hall/CRC, London

Ihaka R and Gentleman R (1996) R: a language for data analysis and graphics. J Comput. Graph Statist. 5:299–314

Schaid DJ (1996) General score statistics for associations of genetic markers with disease using cases and their parents. Genet Epidemiol 13:423–449

Schaid DJ (1999) Likelihoods and TDT for the case-parents design. Genet Epidemiol 16:250–260

Schaid DJ, Sommer SS (1993) Genotype relative risks: methods for design and analysis of candidate-gene association studies. Am J Hum Genet 53:1114–1126

Schaid DJ and Sommer SS (1994) Comparison of statistics for candidate–gene association studies using cases and parents. Am J Hum Genet 55:402–409

Self SG, Longton G, Kopecky KJ, and Liang KY (1991) On estimating HLA–disease association with application to a study of aplastic anemia. Biometrics 47:53–61

Sham PC (1998) Statistics in human genetics. Arnold, London

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