The triangle test statistic (TTS): a test of genetic homogeneity using departure from the triangle constraints in IBD distribution among affected sib-pairs

M.H. Dizier, Hadi Quesneville, B. Prum, H. Selinger-Leneman, F. Clerget-Darpoux

Ann. Hum. Genet. (2000), **64**, 433–442
Printed in Great Britain

The triangle test statistic (TTS): a test of genetic homogeneity using departure from the triangle constraints in IBD distribution among affected sib-pairs

M. H. DIZIER[1], H. QUESNEVILLE[1], B. PRUM[2], H. SELINGER-LENEMAN[1] AND F. CLERGET-DARPOUX[1]

[1] *INSERM U535, Kremlin Bicêtre, France*
[2] *La Genopole, Université d'Evry, Evry, France*

(Received 5.1.00. Accepted 5.6.00)

SUMMARY

The proportions of affected sibs sharing 2, 1 or 0 identical by descent parental marker alleles have been shown to conform to the 'triangle constraints' (Suarez, 1978; Holmans, 1993). It has also been shown (Dudoit & Speed, 1999) that the constraints are verified provided certain assumptions hold. In this study we explore a realistic situation in which the constraints fail due to the presence of a factor in which the sibs differ, a factor on which penetrance depends. This factor may be a characteristic of the trait (severe vs. mild form), or the presence/absence of an associated trait or an environmental factor. We show that under such situations, using the triangle constraints may lead to important loss of power to detect linkage by the MLS test. We propose here an alternative approach in order to detect both linkage and heterogeneity.

INTRODUCTION

Sib-pair methods are standard tools for linkage studies in complex genetic diseases. The principle of these methods is to assess the sharing of marker alleles identical by descent (IBD) between affected sib-pairs and to conclude in favour of linkage if the observed IBD distribution differs from that expected under the hypothesis of independence of segregation of the disease and the markers. This comparison can be made with a model-free method of linkage analysis, the MLS test (Risch, 1990). Since the IBD value between sibs may be ambiguous in some situations, as in the case of homozygous or unknown parental genotypes, a maximum likelihood method is used to estimate the proportions of sib-pairs IBD = 2, 1 or 0 (respectively z2, z1 and z0) given the genotype observations. Absence of linkage is then tested by the maximum likelihood ratio statistic $\log(L(Z)/L(Z_0))$, with Z the estimated vector of proportions (z2, z1, z0) and Z_0 the vector expected under the hypothesis of genetic independence ($Z_0 = 0.25, 0.5, 0.25$). It has been shown that under the assumptions of Hardy–Weinberg equilibrium at the susceptibility locus, the proportions z2, z1 and z0 among affected sib-pairs are constrained by: $2z0 \leqslant z1 \leqslant 0.5$ (Suarez, 1978; Holmans, 1993). Dudoit & Speed (1999) then showed that in the case of monotonicity of penetrances, triangle constraints were verified without the assumption of random mating of the Hardy–Weinberg equilibrium hypothesis, while these assumptions remained necessary in the case of, for example, overdominance. The validity of the triangle constraints has also been verified for two-locus models under the assumption of linkage equilibrium between the two disease loci (Cordell *et al.* 1995). Holmans (1993) proposed to take these 'triangle constraints' into account in maximizing the MLS and showed that it improved the test's power to detect linkage.

Correspondence: Marie-Hélène Dizier, INSERM U.535, Bâtiment INSERM Gregory Pincus (Rez-de-Chaussée), 80 rue du Général Leclerc, 94276 Le Kremlin Bicêtre Cedex, France. Tel: (33) 1 49 59 53 45; Fax: (33) 1 49 59 53 31. E-mail: dizier@kb.inserm.fr

Then, assuming Hardy–Weinberg equilibrium at the disease locus, the triangle constraints are valid as long as the phenotypes of the two sibs in all pairs are determined by the same genetic model (i.e. as long as the probability for a given genotype to be affected is the same for both). In this study, we present many realistic situations where the triangle constraints do not apply. These situations can occur when the two sibs differ for a factor that modifies the penetrance function. This factor may be a specificity of the trait itself, e.g. severe vs. mild form, or the presence/absence of an associated trait or an environmental factor. For such situations, we propose a new strategy that uses departure from the triangle constraints to search for both linkage and heterogeneity.

Departure from the triangle constraints

Models

Let us consider a phenotype A and a factor B. Let there be a bi-allelic locus H (H,h) at which Hardy–Weinberg equilibrium is assumed. This gene is involved in trait A, but with an effect that differs according to whether or not factor B is present. The parameters of this model are q, the frequency of the allele h, and f_{AB+} and f_{AB-}, the probabilities of the phenotype A given the genotype and the presence or not of factor B, such that:

$$\begin{array}{rccc} & hh & hH & HH \\ f_{AB+} = & (f1, & f2, & f3) \\ f_{AB-} = & (f1', & f2', & f3'). \end{array}$$

IBD distribution calculated under genetic models

Consider a sample of sib-pairs all of whom have trait A: among them, a proportion β are discordant for B. Since it will be assumed later that B is always present in at least one sib of the pair, the proportion of concordant pairs for B is simply $(1-\beta)$.

(1) *When all pairs are discordant for B* ($\beta = 100\%$). The IBD distribution expected at a highly polymorphic (heterozygosity equal to 100%) marker M strictly linked to the locus H (negligible recombination fraction), among sib pairs concordant for A but discordant for B, has been calculated under different models. We calculated the IBD distribution, first under simple models with $q = 0.1$, $f3 = f3' = 0$, $f1 = f1' = 0.1$ and penetrance ratio values λ (= f2/f1) and λ' (= f2′/f1′) ranging from 0 to 5.

Table 1 reports the IBD distribution for these models. Departure from the triangle constraints occurs for the models studied here, only when the order of the penetrance vectors f_{AB+} and f_{AB-} is inverted (i.e. $\lambda > 1$ and $\lambda' < 1$ or vice versa). The larger the difference between λ and λ', the greater the departure from the triangle constraints. For example, consider the model with $\lambda = 0$ and $\lambda' = 5$: the expected vector $(z2, z1, z0)$ is $(0.09, 0.83, 0.08)$ and thus deviates substantially from the constraint $z1 \leqslant 0.5$. When $\lambda = 0$ and $\lambda' = 1$ (corresponding respectively, to a recessive and a dominant model), the vector (z2,z1,z0) is equal to $(0.31, 0.63, 0.06)$. This departure from the triangle constraint $z1 \leqslant 0.5$, although smaller than in the preceding example, is nonetheless substantial.

This type of model (where $\lambda > 1$ and $\lambda' < 1$ or vice versa) may correspond, for example, to a disease whose expression is severe in subjects homozygous for the 'disease' allele and mild in those who are heterozygous. Another example is the case of two traits A and B, the association of which is assumed to be due only to a locus H involved in both diseases. If trait A is determined at locus H by a dominant genetic model with a penetrance vector equal to $f_A = (0.9, 0.9, 0.10)$ and trait B by a recessive model with $f_B = (0.9, 0.01, 0.01)$, the resulting penetrance vector for the phenotype of A

Table 1. *Distribution IBD (vector Z) expected among sib-pairs concordant for trait A but discordant for factor B, for various values of λ and λ′*

$\lambda' = 5$								0.45 0.47 0.08
$\lambda' = 4$							0.45 0.47 0.08	0.45 0.47 0.08
$\lambda' = 3$						0.45 0.47 0.08	0.45 0.47 0.08	0.45 0.47 0.08
$\lambda' = 2$					0.44 0.48 0.08	0.44 0.48 0.08	0.44 0.47 0.08	0.44 0.47 0.08
$\lambda' = 1$				0.43 0.49 0.08	0.43 0.49 0.08	0.43 0.48 0.08	0.43 0.48 0.08	0.43 0.48 0.08
$\lambda' = 0.5$			0.42 0.5 0.08	0.42 0.5 0.08	0.42 0.5 0.08	0.42 0.5 0.08	0.42 0.5 0.08	0.42 0.5 0.08
$\lambda' = 0.1$		0.59 0.37 0.04	0.43 0.5 0.06	**0.38 0.54 0.07**	**0.35 0.57 0.08**	**0.33 0.58 0.08**	**0.33 0.59 0.08**	**0.32 0.59 0.08**
$\lambda' = 0$	0.83 0.17 0.0**	0.71 0.27 0.02	0.5 0.45 0.05	**0.31 0.63 0.06**	**0.19 0.74 0.07**	**0.14 0.78 0.08**	**0.11 0.81 0.08**	**0.09 0.83 0.08**
	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$

λ, λ' are the respective penetrance ratios (f2/f1) and (f2′/f1′) of the penetrance vectors f_{AB+} = (f1, f2, f3) and f_{AB-} = (f1′, f2′, f3′)
IBD distributions have been calculated with q = 0.1, f1 = f1‘ = 0.1 and f3 = f3′ = 0
Vector Z = z2 z1 z0 defined in the text.
Results are symmetrical according to λ and λ'.
Z vectors written in bold type correspond to the case where $z_1 > 0.5$.

Table 2. *Expected IBD distribution among sib-pairs discordant for factor B under different genetic models*

Models	Z vector (z2, z1, z0)	Area
M1 $q = 0.3$, $f_A = (0.9, 0.9, 0.1)$ $f_B = (0.9, 0.01, 0.01)$ $(f_{AB+}) = (0.81, 0.009, 0.001)$ $(f_{AB-}) = (0.09, 0.891, 0.099)$	0.07, 0.70, 0.23	1
M2 $q = 0.5$, $f_A = (1, 1, 0.01)$, $f_B = (0.9, 0.009, 0.009)$ $(f_{AB+}) = (0.9, 0.009, 0.00009)$ $(f_{AB-}) = (0.1, 0.991, 0.00991)$	0.07, 0.63, 0.30	2
M3 $q = 0.01$, $f_A = (1, 0.01, 0.01)$, $f_B = (0.9, 0.9, 0.009)$ $(f_{AB+}) = (0.9, 0.009, 0.00009)$ $(f_{AB-}) = (0.1, 0.001, 0.00991)$	0.57, 0.24, 0.19	3
M4 $q = 0.1$, $f_A = (1, 0.1, 0.1)$, $f_B = (0.9, 0.9, 0.009)$ $(f_{AB+}) = (0.9, 0.09, 0.0009)$ $(f_{AB-}) = (0.1, 0.01, 0.0991)$	0.21, 0.40, 0.39	4
M5 $q = 0.5$, $f_A = (0.5, 0.3, 0.1)$, $f_B = (0.9, 0, 0)$ $(f_{AB+}) = (0.45, 0,0)$ $(f_{AB-}) = (0.05, 0.3, 0.1)$	0.09, 0.60, 0.32	5

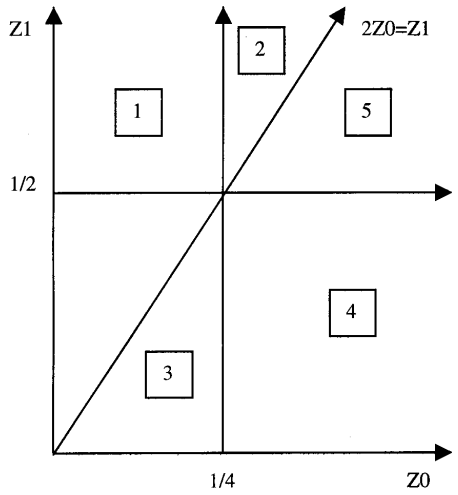

Fig. 1. Division of the (Z0, Z1) plane into areas (1–5) when there is departure from the triangle constraints.

and B together (f_{AB+}) is then equal to (0.81, 0.009, 0.001), and the vector for the phenotype of A without B (f_{AB-}) is (0.09, 0.891, 0.099). When B is present, the homozygous genotype confers the highest risk; conversely, when B is absent, the heterozygous genotype is at the greatest risk.

All the models we have so far considered lead to departure from the constraint $z1 \leqslant 0.5$. There are also models that lead to departure from the constraints $2z0 \leqslant z1$, and $z0 \leqslant 0.25$, which is a consequent constraint upon the two preceding ones. Table 2 includes some simple associated-disease models M_i (Table 2) in which the IBD distribution falls outside the triangle into areas A_i, depending on which constraint(s) is (are) not respected (Fig. 1). In every model leading to a departure from the triangle constraints we observe an inversion in the penetrance rank order between the two vectors f_{AB+} and f_{AB-}. However, no simple rules appear to predict which constraint will be violated.

Note, however, that when all penetrances are equal in one of the two vectors f_{AB+} or f_{AB-}, i.e. when the locus H has no effect on the corresponding phenotype, the IBD distribution expected in

Table 3. *IBD distribution expected for various proportions β of discordant pairs under the generating model of two associated diseases A and B (f_{AB+} = (0·81,0·009,0·001), f_{AB-} = (0·09, 0·891, 0·099) and $q = 0{\cdot}1$)*

β	z2	z1	z0
0.0	0.79	0.19	0.01
0.1	0.73	0.25	0.02
0.2	0.66	0.31	0.04
0.3	0.59	0.36	0.05
0.4	0.52	0.42	0.06
0.5	0.45	0.48	0.07
0.6	**0.39**	**0.53**	**0.08**
0.7	**0.32**	**0.59**	**0.09**
0.8	**0.25**	**0.64**	**0.11**
0.9	**0.18**	**0.70**	**0.12**
1.	**0.08**	**0.79**	**0.13**

Lines in bold type correspond to IBD distributions with some departure from triangle constraints.

discordant pairs is (0.25, 0.5, 0.25). Consequently, departure from the constraints implies that the genetic factor detected is always involved in A but that its effect differs according to whether B is present or not.

(2) *For various proportions β between 0 and 1.* Most often, any group of sib-pairs selected for disease A will contain a mixture of pairs concordant and discordant for B. In other words, the proportion β of discordant pairs may vary between 0 and 1, depending on the frequency of B and on the strength of association between A and B. Some types of ascertainment may increase the proportion of discordant pairs. In particular, some selection modes imply B for the proband, the first sib selected, but have no particular implication about B and the second sib. One example of this relatively frequent situation is hospital-based ascertainment: the proband will probably have a severe form of the disease, while the sib may not.

Let us now consider this type of situation, where B is present in the first sib of all sib-pairs, but is present in the second sib only in a proportion $(1-\beta)$ of pairs. We calculated the IBD distribution expected at a marker M strictly linked to the locus H under the model of two associated diseases (A and B) with f_A = (0.9, 0.9, 0.1) and f_B = (0.9, 0.01, 0.01) corresponding to the penetrance vectors f_{AB+} = (0.81, 0.009, 0.001) and f_{AB-} = (0.09, 0.891, 0.099), with q = 0.1, and for different values of β ranging between 0 and 1. The IBD distributions calculated here are the mixture of the IBD distributions among pairs concordant for B (among which the triangle constraints are expected to be verified) and among pairs discordant for B.

The IBD distributions, reported in Table 3, show that values of β greater than 0.5 lead to departure from the triangle constraints. This proportion of discordant pairs is not unrealistic since the expected proportion β of discordant pairs is 0.71 under the model and mode of ascertainment described above. For this value of β, the expected IBD distribution is (0.32, 0.59, 0.09), and the departure from the triangle constraints is not negligible.

Using the MLS to detect linkage in the presence of heterogeneity, with and without the triangle constraints

It has been shown that in absence of heterogeneity, introduction of the triangle constraints in the maximization of the MLS increased the power to detect linkage (Holmans, 1993). It would thus be interesting to examine the influence of the triangle constraints on the power of the MLS test to detect

Table 4. *Mean number of sib-pairs required to detect linkage with type 1 error of 1‰ and 50% power with MLS_c versus MLS_u (N_{MLSc}/N_{MLSu}), estimated respectively, with and without the constraints*

Models*:	M1	M2	M3	M4	M5
$\beta = 1$	∞/**54**	∞/**62**	24/29	∞/**145**	∞/**80**
$\beta = 0.9$	∞/**103**	∞/**97**	20/24	∞/**242**	∞/**131**
$\beta = 0.8$	∞/**195**	∞/**189**	16/20	**2283/323**	∞/**278**
$\beta = 0.7$	**2667/313**	∞/**444**	14/17	**430/275**	∞/**667**
$\beta = 0.6$	**374/303**	∞/**1132**	12/15	180/201	∞/**2727**
$\beta = 0.5$	169/208	**3053/1395**	10/13	103/130	**3200/2804**
$\beta = 0.4$	89/115	510/612	#/11	67/87	430/552
$\beta = 0.3$	51/60	205/265	#/10	46/60	184/238
$\beta = 0.2$	34/44	119/154	#/#	32/41	108/140
$\beta = 0.1$	24/31	72/93	#/#	23/30	67/87
$\beta = 0$	19/24	48/62	#/#	17/21	48/62

*, defined in Table 2.
β, proportion of discordant pairs.
∞, > 10000 sib-pairs.
#, < 10 sib-pairs.
Results in bold type correspond to the case where the mean number of sib-pairs required by the MLS_c is greater than the one required by the MLS_u.

linkage in the presence of heterogeneity. We considered the 5 different genetic models in Table 2, which, in discordant sib-pairs, result in IBD distribution in each of the 5 areas outside the triangle (Fig. 1). We also consider various values for β, the proportion of discordant pairs. We calculated the sample sizes required to detect linkage with each statistic, MLS_c (estimated with constraints) and MLS_u (estimated without constraints), using the exact IBD distributions expected under given models in a sample of infinite size. These exact distributions are calculated assuming a complete polymorphic marker and known parental genotypes at this marker

Let $Z_u = (Z0_u, Z1_u, Z2_u)$ be the IBD distribution expected under the generating model, $Z_c = (Z0_c, Z1_c, Z2_c)$ be the IBD distribution expected under the same model but estimated with the triangle constraints, and Z_0 be the distribution under the hypothesis of no linkage $(0.25, 0.5, 0.25)$. For a sample size N, the expected MLS_c score can be deduced as follows:

$$MLS_c = \log(L(Z_c)/L(Z_0)) = \log\left(Z2_c^{N.Z2u}.Z1_c^{N.Z1u}.Z0_c^{N.Z0u}/0.25^{N.Z2u}.0.5^{N.Z1u}.0.25^{N.Z0u}\right)$$

$$= N.[Z2_u.\log(Z2_c/0.25) + Z1_u.\log(Z1_c/0.5) + Z0_u.\log(Z0_c/0.25)].$$

Similarly the MLS_u score can be calculated by replacing Z_c with Z_u.

The sample sizes N for which the MLS_c and MLS_u values exceed a given threshold may be calculated. The power corresponding to these sizes is roughly 50%. For a type 1 error of $1°/_{oo}$ and a highly polymorphic marker, the threshold is 2.32 for MLS_c (given by Holmans, 1993) and 3 for MLS_u (this threshold can be determined easily given that $2\ln(10)$; MLS_u follows a χ^2 distribution asymptotically with 2 D.F.). Table 4 reports sample sizes required for a type 1 error of $1°/_{oo}$ and 50% power, for the five models and various values of β. For models M2 and M5, when the proportion of discordant pairs is 0.5 or more, the MLS_u requires much smaller sample sizes than the MLS_c test. The same is true for models M1 and M4 when the β values equal or exceed 0.6 and 0.7, respectively. Moreover, for some proportions β in these models (exceeding a value that ranges from 0.5 to 0.8), linkage can never be detected with the MLS_c test: when the vector Z is estimated within the triangle constraints, it tends to be very close to the null hypothesis vector (0.25, 0.5, 0.25), so that a sample

of an infinite number of pairs is required for the MLS_c. In contrast, very reasonable sample sizes will be sufficient to detect linkage with the MLS_u test. On the other hand, when β is less than 0.5, the MLS_u test becomes less powerful than the MLS_c test for the models considered here. In the absence of heterogeneity (i.e. $\beta = 0$), however, the differences between the sample sizes required for each method do not exceed 15 pairs.

For model M3, the MLS_c test can detect linkage with small sample sizes even when the proportion of discordant pairs is 100%. However, the sizes required by the MLS_u are only slightly larger (differences of size no more than 5 pairs).

It appears that for some models, the gain in power obtained by using the triangle constraints in the absence of heterogeneity is much smaller than the loss of power that occurs when these constraints are used in the presence of heterogeneity.

Strategy of detection of linkage accounting for heterogeneity

When heterogeneity is suspected, it may thus be appropriate to use the MLS_u rather than the MLS_c, since the former allows linkage to be detected in either the presence or absence of heterogeneity.

We propose the following strategy, which includes an initial search for linkage with the MLS_u test. The MLS_u score can be transformed by $2.\mathrm{Ln}(10).MLS_u = 2\mathrm{Ln}\ [L(Z_u)/L(Z_0)]$, which follows a χ^2 distribution asymptotically with 2 D.F. If linkage is detected by this first test, we propose a second step to search for heterogeneity by testing the triangle constraints. The triangle test statistic (TTS) compares the vector Z estimated without constraints (Z_u) with the vector estimated with constraints (Z_c). It is calculated as follows: $TTS = \log\{L(Z_u)/L(Z_c)\}$. The null hypothesis tested is linkage with genetic homogeneity. Rejection of the triangle constraints would thus allow to conclude for heterogeneity, i.e. that some sib-pairs are discordant for the presence of a factor that modifies the effect of the susceptibility gene. However, given the other assumptions required for the triangle test, such a conclusion would hold conditionally on Hardy–Weinberg equilibrium at the susceptibility locus.

Note that the TTS is also equivalent to the difference between the MLS values estimated with and without constraints. As Holmans (1993) did for the MLS distribution, we used the method of Self & Liang (1987) to calculate the asymptotic distribution of TTS as a mixture of $\chi^2_{1\mathrm{df}}$ and $\chi^2_{2\mathrm{df}}$, subject to the condition that, in a first step, the MLS_u test yielded a value greater than a threshold A chosen as the criteria for a conclusion of linkage (see Appendix I). We calculated and presented in Table 5 the criteria for the TTS test of any size and considering various values of A. Note that to apply the TTS test conditionally on MLS_u exceeding a given threshold A leads to an increase in the criteria of the TTS, in comparison with the situation where the TTS would be directly applied. However, the interest of this strategy in two steps is first to detect linkage (in both the presence or the absence of heterogeneity) and second to search for such heterogeneity. Since the TTS distribution was derived under the hypothesis of no linkage, this test is quite conservative. The values are calculated with a program written in S language that is available on request from the authors.

Note that in the case of genome screening, the strategy proposed here requires some correction for multiple testing. The initial analyses with the MLS_u test can be corrected as in other linkage studies, either by using thresholds like those calculated by Krugklyak & Lander (1995) on the assumption of a very dense and polymorphic map, or by calculating the p values corresponding to the real marker map, by simulation. The second step of the analysis is conditional on the first step and involves only

Table 5. *Thresholds of various sizes for the TTS test conditionally on MLS_u exceeding a given value A*

(a) When parents are typed

		Size of test				
A	nb alleles	0.05	0.01	0.001	0.0001	0.00001
2	2	3.02	3.71	4.70	5.69	6.69
	5	3.01	3.70	4.69	5.69	6.68
	20	3.01	3.70	4.69	5.69	6.68
2.2[a]	2	3.22	3.91	4.90	5.89	6.89
	5	3.21	3.90	4.89	5.89	6.88
	20	3.21	3.90	4.89	5.89	6.88
3	2	4.01	4.70	5.69	6.69	7.68
	5	4.00	4.69	5.69	6.68	7.68
	20	4.00	4.69	5.68	6.68	7.67
3.6[b]	2	4.59	5.29	6.29	7.28	8.28
	5	4.59	5.28	6.28	7.28	8.27
	20	4.58	5.28	6.28	7.28	8.27
4	2	5.00	5.69	6.69	7.68	8.68
	5	4.99	5.69	6.68	7.68	8.67
	20	4.99	5.69	6.68	7.67	8.67
(b) When parents are not typed						
2	2	3.04	3.73	4.72	5.72	6.71
	5	3.03	3.72	4.71	5.70	6.70
	20	3.01	3.70	4.69	5.69	6.68
2.2[a]	2	3.24	3.93	4.92	5.91	6.91
	5	3.23	3.92	4.91	5.90	6.90
	20	3.21	3.90	4.89	5.89	6.88
3	2	4.03	4.72	5.72	6.71	7.71
	5	4.02	4.71	5.70	6.70	7.69
	20	3.99	4.69	5.69	6.68	7.68
3.6[b]	2	4.62	5.31	6.31	7.31	8.31
	5	4.60	5.30	6.30	7.30	8.29
	20	4.59	5.28	6.28	7.28	8.27
4	2	5.02	5.72	6.71	7.71	8.70
	5	5.01	5.70	6.70	7.69	8.69
	20	4.99	5.69	6.69	7.68	8.67

[a] Threshold corresponding to $p = 7.10^{-4}$ proposed by Lander & Kruglyak (1995) as suggestive linkage.
[b] Threshold corresponding to $p = 2.10^{-5}$ proposed by Lander & Kruglyak (1995) as significant linkage.

the markers detected by the MLS_u test. It can thus be, for simplicity's sake, corrected by the number of tested markers with the Bonferroni correction.

Discussion

Most complex diseases are presented over a broad clinical spectrum (e.g. rheumatoid arthritis, with or without rheumatoid factor, subcutaneous nodules or other extra-articular manifestations). For these diseases, the phenotype–genotype correspondence is not often evident; one difficulty is classifying individuals as affected or unaffected. The issue of classification has often been raised in linkage analysis as a problem that can lead to a substantial decrease of the power to detect linkage. Moreover, for complex diseases, there may be a large number of associated traits that are not specific to the disease under study (e.g. asthma with allergy, bronchial hyperresponsiveness) and environmental factors (allergy with smoking habits, pollution, animals, pollen exposure). Similarly, these diseases may depend on several genes. All these factors may interact with the genes involved in a disease. In such cases, the same genetic model would not underlie the phenotypes of two

individuals affected by the same disease but differing for the presence of one of these associated factors. We show here that such heterogeneity between sibs may induce departure from the triangle constraints in the IBD distribution and substantially decrease the power to detect linkage when using the constraints of the MLS_c test. Under some models of heterogeneity, there may be very large differences in power depending upon whether the triangle constraints are applied or not, i.e. on whether the MLS_c or the MLS_u test is used. The advantage of the strategy we propose here, using the MLS_u followed by TTS, rather than the MLS_c, will clearly depend on the underlying model and on the proportion of pairs discordant for this factor. This proportion depends on the ascertainment mode, but we show here realistic situations that can result in a large proportion of discordant pairs. One such example is the situation where the selection criteria imply that the first selected sib (proband) has B (or does not have B). An interesting illustration is the application of the TTS test to simulated data of sib-pairs affected by a disease having two forms (severe or mild), which were provided by GAW 11 (Quesneville *et al*. 1999). Since the ascertainment mode required that the first sib have a severe form of the disease, the proportion of pairs discordant for disease severity was quite large. The TTS test led to the conclusion that a genetic factor linked to one of the markers studied had a different type of involvement in the severe and mild forms of the disease.

In conclusion, the gain in power obtained by using the triangle constraints with the MLS_c test in the absence of heterogeneity appears smaller than the loss of power that occurs when these constraints are used under some models of heterogeneity. Moreover, for most complex diseases it is not possible to control concordance for all associated factors when selecting affected sib-pairs, because of the number of factors that may be present. There may also be unidentified factors, which could not in any case be controlled for. In such situations, the strategy proposed here, the MLS_u test followed by the TTS, may be of interest because it would allow linkage to be detected even in the presence of possible heterogeneity and would, moreover, detect such heterogeneity.

References

Cordell, H. J., Todd, J. A., Bennett, S. T., Kawaguchi, Y. & Farrall, M. (1995). Two-locus maximum lod score analysis of a multifactorial trait: joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes. *Am. J. Hum. Genet.* **57**, 920–934.

Dudoit, S. & Speed, T. P. (1999). Triangle constraints for sb-pair identity by descent probabilities under general multilocus model for disease susceptibility. In *Statistics in the Health Sciences: Genetics* (ed. S. Geisser & E. A. Halloran). IMA Volumes in Mathematics and its Applications, Springer Verlag.

Holmans, P. (1993). Asymptotic properties of affected-sib-pair linkage analysis. *Am J. Hum. Genet.* **52**, 362–374.

Kruglyak, L. & Lander, E. S. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**, 241–247.

Morton, N. E. (1956). The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am. J. Hum. Genet.* **8**, 80–96.

Quesneville, H., Dizier, M. H. & Clerget-Darpoux, F. (1999). Departure from the triangle constraints in discordant sib-pairs: a test for genetic heterogeneity. *Genet. Epidemiol.* **17**, S685-S689.

Risch, N. (1990). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* **46**, 229–241.

Suarez, BK. (1978). The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens* **12**, 87–93.

Self, S. G. & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J. Am. Stat. Assoc.* **82**, 605–610.

Appendix I

Derivation of the asymptotic distribution of the triangle test statistic (*TTS*)

Because the TTS test will be applied to test genetic homogeneity only on markers for which linkage has been concluded by the MLS_u test without constraints, the distribution of TTS is derived conditionally on MLS_u exceeding a given threshold A, the criterion used to conclude for linkage.

The conditional probability that TTS exceeds a value K is:

$$\begin{aligned}\alpha &= \mathrm{P}(\mathrm{TTS\ K} \mid \mathrm{MLS_u} > \mathrm{A}) \\ &= \mathrm{P}(\mathrm{MLS_u} - \mathrm{MLS_c} > \mathrm{K} \text{ and } \mathrm{MLS_u} > \mathrm{A})/\mathrm{P}(\mathrm{MLS_u} > \mathrm{A}) \\ &= \mathrm{P}(\mathrm{MLS_u} > \sup(\mathrm{A}, \mathrm{MLS_c} + \mathrm{K}))/\mathrm{P}(\mathrm{MLS_u} > \mathrm{A}) \qquad (1)\end{aligned}$$

with $\mathrm{P}(\mathrm{MLS_u} > \mathrm{A}) = \mathrm{P}(\chi^2_{2\mathrm{df}} > \mathrm{A})$

As in Holmans (1993), we derived the numerator γ of (1) for the four regions V, T‘, U1, U2 in the plane of the normalized transform (z0, z1) of (Z0, Z1):

$\gamma = \Sigma_i\ \gamma_i = \Sigma_i \mathrm{P}(\mathrm{MLS_u} \in \mathrm{i})\ \mathrm{P}(\mathrm{MLS_u} > \sup(\mathrm{A}, \mathrm{MLS_c} + \mathrm{K}) \mid \mathrm{MLS_u} \in \mathrm{i})$

1. For $\mathrm{MLS_u} \in \mathrm{T}' : \gamma_1 = 0$
2. For $\mathrm{MLS_u} \in \mathrm{V} : \gamma_2 = [(\pi - \theta)/2\pi]\ \mathrm{P}(\mathrm{MLS_u} > \sup(\mathrm{A}, \mathrm{K})) = [(\pi - \theta)/2\pi]\ \mathrm{P}(\chi^2_{2\mathrm{df}} > \sup(\mathrm{A}, \mathrm{K}))$
3. For $\mathrm{MLS_u} \in$ U1 or U2:

if $\mathrm{K} < \mathrm{A}$ then

$$\begin{aligned}\gamma_3 &= 1/4[\textstyle\int_0^{\mathrm{A}-\mathrm{K}} \mathrm{P}(\mathrm{MLS_u} - \mathrm{MLS_c} > \mathrm{A} - \mathrm{y})\mathrm{P}(\mathrm{MLS_c} = \mathrm{y})\mathrm{dy} \\ &\quad + \textstyle\int_{\mathrm{A}-\mathrm{K}}^{\infty} \mathrm{P}(\mathrm{MLS_u} - \mathrm{MLS_c} > \mathrm{K})\mathrm{P}(\mathrm{MLS_c} = \mathrm{y})\mathrm{dy}] \\ &= 1/4[\textstyle\int_0^{\mathrm{A}-\mathrm{K}} \mathrm{P}(\chi^2_{1\mathrm{df}} > \mathrm{A} - \mathrm{y})\mathrm{P}(\chi^2_{1\mathrm{df}} = \mathrm{y})\mathrm{dy} + \mathrm{P}(\chi^2_{1\mathrm{df}} > \mathrm{K})\mathrm{P}(\chi^2_{1\mathrm{df}} > \mathrm{A} - \mathrm{K})]\end{aligned}$$

if $\mathrm{K} > \mathrm{A}$ then

$$\gamma_3 = 1/4\mathrm{P}(\mathrm{MLS_u} - \mathrm{MLS_c} > \mathrm{K}) = 1/4\,\mathrm{P}(\chi^2_{1\mathrm{df}} > \mathrm{K})$$

with $\theta/2\pi$ being the probability to be in the triangle (T′).