Badner JA, Goldin LR (1997) Bipolar disorder and chromosome 18: an analysis of multiple data sets. Genet Epidemiol 14:569–574

———. Meta-analysis of linkage studies. Genet Epidemiol (in press)

Baron M (1999) Optimal ascertainment strategies to detect linkage to common disease alleles. Am J Hum Genet 64: 000–000 (in this issue)

Berrettini WH, Ferraro TN, Goldin LR, Detera-Wadleigh SD, Choi H, Muniec D, Guroff JJ, et al (1997) A linkage study of bipolar illness. Arch Gen Psychiatry 54:27–35

Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI Jr, Gershon ES (1994) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Natl Acad Sci USA 91: 5918–5921

Detera-Wadleigh SD, Badner JA, Goldin LR, Berrettini WH, Sanders AR, Rollins DY, Turner G, et al (1996) Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder, on 21q. Am J Hum Genet 58:1279–1285

Detera-Wadleigh SD, Badner JA, Yoshikawa T, Sanders AR, Goldin LR, Turner G, Rollins DY, et al (1997) Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 4, 7, 9, 18, 19, 20, and 21q. Am J Med Genet 74:254–262

Knowles JA, Rao PA, Cox-Matise T, Loth JE, de Jesus GM, Levine L, Das K, et al (1998) No evidence for significant linkage between bipolar affective disorder and chromosome 18 pericentromeric markers in a large series of multiplex extended pedigrees. Am J Hum Genet 62:916–924

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

Rice J (1997) Genetic analysis of bipolar disorder: summary of GAW10. Genet Epidemiol 14:549–561

Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, et al (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57:1384–1394

Address for correspondence and reprints: Dr. Judith Badner, Department of Psychiatry, University of Chicago, 5841 South Maryland Avenue, MC 3077, Chicago, IL, 60637. E-mail: jbadner@yoda.bsd.uchicago.edu

0002-9297/99/6404-0044$02.00

Am. J. Hum. Genet. 64:1248–1252, 1999

Down-Weighting of Multiple Affected Sib Pairs Leads to Biased Likelihood-Ratio Tests, under the Assumption of No Linkage

To the Editor:

In large nuclear families with several affected siblings, affected-sib-pair analyses are often based on all the possible pairs of affected siblings that can be formed. Because of concern about dependence between sibling pairs from the same family, it has been recommended that, when tests for linkage are constructed, pairs from families with a large number of affected siblings be given less weight (e.g., see Daly and Lander 1996; Davis and Weeks 1997). Various weighting schemes have been proposed (Hodge 1984; Suarez and Van Eerdewegh 1984; Sham et al. 1997). However, several authors have shown in simulations that likelihood-ratio tests weighted according to the proposal of Suarez and Van Eerdewegh (1984) are quite conservative (Meunier et al. 1997; Abel and Müller-Myhsok 1998). Likelihood-ratio tests behave differently from other tests for linkage in sibling pairs, such as the means test, and therefore should be treated differently.

It is important to distinguish between the performance of a test statistic under the null hypothesis of no linkage (type I error) and the performance when there is linkage (power). Sham et al. (1997), Blackwelder and Elston (1985), and Suarez and Van Eerdewegh (1984) have shown that, for any weighting function, the means test statistic is unbiased and that, for large numbers of families, the test has the expected type I error under the null hypothesis of no linkage. Different weighting functions can, however, improve the power of the linkage tests. Sham et al. (1997) also have shown that the most powerful weighting function for the means test will be a function of the true genetic model.

However, linkage tests based on the likelihood ratio, such as the maximum-likelihood score (MLS) described by Risch (1990), Holmans (1993), and Kruglyak and Lander (1995), can have a biased distribution (i.e., one that is not χ^2) when there are multiple affected sib pairs per family, even under the null hypothesis. Unweighted likelihood-ratio tests can be slightly anticonservative (Kong et al. 1997; Abel and Müller-Myhsok 1998), and, as noted elsewhere, the commonly used weights of Suarez and Van Eerdewegh (1984), which weight each pair by $2/k$, where k is the number of affected siblings, can be extremely conservative.

The deviation from the expected distribution for likelihood-ratio tests under the null hypothesis is due to the distribution of the identity by descent (IBD) scores from multiple sibling pairs. Although it has been shown that the IBD statuses of any two sib pairs from the same family are independent (Suarez and Hodge 1979; Hodge 1984; Blackwelder and Elston 1985), the pairs are not jointly independent (Kong et al. 1997). Also, the distribution of IBD sharing in large sibships will be highly skewed. When small numbers of sibships are analyzed by the means test, the skewness of the IBD distribution can lead to P values that are too small (Kong et al. 1997; Sham et al. 1997), but the impact of skewness decreases as the sample size increases and the central-limit theorem

applies. Furthermore, when the IBD status is not known with certainty, because of either incomplete marker information or missing parents, then there will be correlation between the estimated IBD proportions from different sib pairs within a family (Kong et al. 1997), which can bias tests of significance, even for large numbers of families. For likelihood-ratio tests, however, we show that the conventional weights of Suarez and Van Eerdewegh (1984) produce tests that, although conservative, deviate more strongly from the expected distribution than do the unweighted tests, even when a large number of families are studied and the IBD status is known with certainty.

The likelihood ratio–test statistic is 2 $\log_e$ of the ratio of the likelihood of sib-pair data when the three expected allele-sharing proportions are estimated—$L(\hat{z}_0,\hat{z}_1,\hat{z}_2)$—to the likelihood when these proportions are fixed at their null-hypothesis values of (.25, .5, .2.5). When no constraints are put on the estimated proportions, this likelihood-ratio test is expected to be distributed as χ^2_2. The MLS of Holmans (1993), implemented in SPLINK (Holmans and Clayton 1995) and MAPMAKER/SIBS (Kruglyak and Lander 1995), is also of this form, but the allele-sharing proportions are constrained to fall within the "possible triangle" bounded by $z_1 = .5$ and $z_1 = 2z_0$, so the constrained likelihood-ratio test has a distribution that is a mixture of χ^2 distributions with 1 df and 2 df (Holmans 1993).

Unconstrained likelihood ratio tests.—Illustrating the skewed distributions that can arise with weighting, figure 1 shows the distribution of unconstrained likelihood ratio–test statistics from a simulation that included many large sibships in which the fully informative marker was *unlinked* to the disease locus. Individuals carrying one or two copies of a disease-susceptibility gene with population frequency .1 were assumed to have a 75% probability of being affected. The sporadic rate was assumed to be 15%, so that the gene increased disease risk fivefold. Sibship size was generated from a geometric distribution with mean .3 truncated at a maximum of 12 sibs, and 100 families with at least 2 affected siblings were ascertained. The mean number of affected siblings per family was 2.6 (range 2–10; 3% of families had $\geq$6 affected sibs). Five hundred data sets were generated. In figure 1, the allele-sharing estimates were not constrained to fall into the possible triangle (Holmans 1993), and so the expected distribution of these statistics should be χ^2_2. It can be seen that the weighted likelihood-ratio tests, using weights of $2/k$, do not follow the expected distribution but that the unweighted tests do, over most of the distribution. (The top 3% of the unweighted test statistics are larger than expected, which would lead to anticonservative tests, as reported by others [Kong et al. 1997; Abel and Müller-Myhsok 1998]).

Let w_k denote a weight applicable to families with k affected siblings; for example, the weight of $w_k = 2/k$ allows each family to contribute an amount proportional to the number of affected sibs. Let z_j be the expected proportion of sib pairs sharing j alleles IBD ($j = 0,1,2$). Let ρ_{ij} be the probability that sib pair i shares j alleles IBD at a particular location, given their marker data, divided by the null-hypothesis probability that j alleles are shared IBD. The sib-pair likelihood ratio–test statistic can be rewritten to group together all families with the same number of affected sibs,

$$LR = 2\sum_{k=2}^{k_{\max}} w_k \sum_{f\in k}\sum_{i\in f} \log\left(\sum_{j=0}^{2}\rho_{ij}z_j\right) ,$$

where $k_{\max}$ is the largest number of affected sibs in the sample of families, the sum over f denotes summation over all families with k affected siblings, and the sum over i is over all sib pairs that can be formed in family f. By use of a Taylor-series expansion of the likelihood around the null-hypothesis values for z_0 and z_1, for each set of families with the same k, the likelihood ratio can be approximated by

$$LR \approx 2\sum_{k=2}^{k_{\max}} w_k[\Delta_k' V_k^{-1}\Delta_k] , \qquad (1)$$

which is a function of the first derivatives with respect to z_0 and z_1,

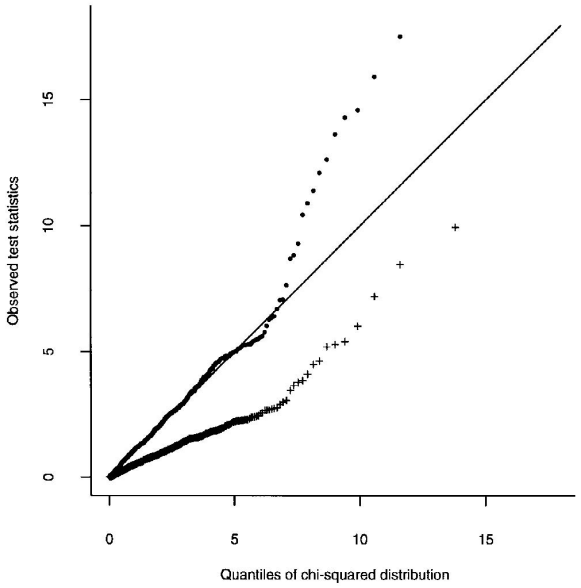

Figure 1 Distribution of weighted and unweighted unconstrained likelihood-ratio tests for linkage, under the null hypothesis of no linkage, for a fully informative marker: dots denote unweighted tests, and plus signs denote weighted tests. The diagonal line shows the expected distribution corresponding to χ^2_2.

$$\Delta_k' = \left[\sum_{f \in k} \sum_{i \in f} \delta_{i0} \qquad \sum_{f \in k} \sum_{i \in f} \delta_{i1} \right],$$

and of the second derivatives,

$$V_k = \begin{bmatrix} \sum_{f \in k} \sum_{i \in f} \delta_{i0}^2 & \sum_{f \in k} \sum_{i \in f} \delta_{i0}\delta_{i1} \\ \sum_{f \in k} \sum_{i \in f} \delta_{i0}\delta_{i1} & \sum_{f \in k} \sum_{i \in f} \delta_{i1}^2 \end{bmatrix},$$

where $\delta_{i0} = (\rho_{i0} - \rho_{i2})$ and $\delta_{i1} = (\rho_{i1} - \rho_{i2})$. The convergence of each of these Taylor-series expansions follows from, for example, the proofs by Self and Liang (1987), as long as there are a sufficient number of families in each category k.

In the means test studied by Sham et al. (1997), the expectation of the numerator of their statistic, $\bar{z} - \frac{1}{2}$, was always 0 under the null hypothesis, for any weighting function. Different weights gave different powers to the test. However, as can be seen from equation (1), the likelihood-ratio test will have different expected values for different weights even under the null hypothesis. The likelihood ratio can be considered as a function of the score variances V_k, rather than as a function of the mean scores. Hence the problem for likelihood-ratio tests is not to choose a weight that gives the best power but to choose a weight that leads to the correct distribution under the null hypothesis. In addition, the likelihood ratio is a measure of the odds in favor of linkage. These ratios will be affected by different choices of weights, and therefore their interpretation becomes difficult.

When sib-pair IBD values are correlated within families, then an appropriate weight, based on the average correlation over a set of families, should correct for this dependence. Suppose that some correlation or clustering does exist between affected sib pairs from the same family. If V is the variance-covariance matrix of the score elements for a set of independent sib pairs, then the desired weight w_k, which should give the correct distribution to the likelihood ratio, is one in which $V^{-1} = w_k V_k^{-1}$. In our simulated data, we estimated the strength of the dependence between multiple sib pairs, to estimate a value for w_k. Obtaining estimates of the best weights for elements of the score vector is similar to the problem of estimating the design effects that are due to two-stage sampling in linear-regression models (Scott and Holt 1982; for details, see Greenwood 1998). By means of the procedure reported by Scott and Holt (1982), the true variance of the estimate of the mean score elements under clustering can be obtained by multiplication of the variance, under simple random sampling, by the design effect, $r_{kj} = 1 + (m_{kj} - 1)\,\zeta_{kj} = \frac{1}{w_k}$, where ζ_{kj} is the intraclass correlation, for element j of Δ_k, between two sib pairs in the same family, and

$$m_{kj} = \frac{\sum_{f \in k} n_k^2 \bar{\delta}_{fj}^2}{\sum_{f \in k} \sum_{i \in f} \delta_{ij}^2},$$

where $\bar{\delta}_{fj}$ is the mean value of δ_{ij} in family f and n_k is the number of sib pairs from a family with k siblings, so that $n_k = k(k-1)/2$.

To evaluate the sizes of design effects for these simulations, the noniterative estimator of Eliasziw and Donner (1991) was used to estimate the intraclass correlation. Let N_k be the number of families in the sample that have k siblings. Then the intraclass correlation can be estimated by

$$\zeta_{kj} = \frac{\sum_{f \in k} v_f \sum_{i \in f} \sum_{i' \neq i} (\delta_{ij} - \bar{\delta}_{kj})(\delta_{i'j} - \bar{\delta}_{kj})}{\sum_{f \in k} v_f (n_k - 1) \sum_{i \in f} (\delta_{ij} - \bar{\delta}_{kj})^2},$$

where $v_f = 1/[n_k(n_k - 1)N_k]$ and $\bar{\delta}_{kj}$ is the sum over pairs and families of the score elements, weighted by $v_f(n_k - 1)$. If this intraclass correlation between the sib pair–score quantities is small, then the value of r_{kj} will be close to 1. Note that the design-effect estimate depends directly on the estimated IBD sharing (ρ_{ij}) and can be calculated before estimation of the allele sharing z_j or the likelihood ratio–test statistic.

The design effect was assumed to be the same for each element of the score vector, $j = 0,1$, so that $r_k = (r_{k0} + r_{k1})/2$. In table 1, the estimated design effects r_k are reported for several different simulations of 1,000 families, under the same generating model as has been used above, for LOD scores based on all pairs of the affected children (where the LOD score is the likelihood-ratio test divided by 2 log[10]). The variability due to sampling can be seen by comparison of the simulations. For families with only two affected children, the intraclass correlation is 0, and the design effect is always 1. The design effects are given for families with three, four, five, or six or more affected children, and the numbers in parentheses are the number of families, per 1,000, that have that number of affected children. In fact, the design effects are almost exactly 1, for a fully informative marker, which is in agreement with Kong et al.'s (1997) statement that the correlation should be 0 for known IBD. However, in the bottom section of table 1, the estimated design effects are reported for another series of simulations, in which IBD was estimated for a marker with four equally frequent alleles, under the assumption that the parents were unavailable. The design effects can be substantially >1, implying positive correlation among the score elements within sibships. Weights <1 would be appropriate for this case, to avoid inflation of the type I error. Notice that, in all cases, the inverse of the $2/k$ weights and the weights based on information content

(Hodge 1984) are all substantially greater than the estimated design effects; use of the conventional weights will produce an overly conservative result.

For unweighted test statistics in figure 1, the deviation from the expected distribution, in the upper tail, is not trivial. Therefore, even for a fully informative marker, the distribution of the likelihood-ratio statistics has not converged to χ^2_2, for 100 families. The divergence gradually decreases with sample size; for 500 families, there is still a visible deviation from the expected distribution (data not shown). This can be contrasted to the good convergence of the means test statistics to the normal distribution for large numbers of families (Sham et al. 1997; Kong et al. 1997). The score elements are negatively skewed in large sibships, and this implies that there tend to be larger deviations (in absolute value) from the mean score when IBD = 2 than when IBD = 0 or 1. Such skewness is consistent with anticonservative likelihood-ratio tests. The design effects estimated above depend on a two-term Taylor-series expansion, and therefore the approximation does not incorporate skewness or higher-order terms. The correct weights for large families will be smaller than those based on these design effects.

Constrained likelihood-ratio tests.—Figure 2 shows the distribution of the constrained likelihood-ratio tests for weighted and unweighted models (when the weights of $2/k$ and the same simulation design are used). It can be seen that the empirical distributions have a pattern similar to that in figure 1. The tail probability, α, for the constrained likelihood-ratio test with independent sibling pairs can be expressed as $\alpha = \frac{1}{2}P(\chi^2_1 > Q) + \frac{\theta}{2\pi}P(\chi^2_2 > Q)$ (Holmans 1993), where the angle θ is a function of the expected information matrix and represents the probability that the allele-sharing estimates

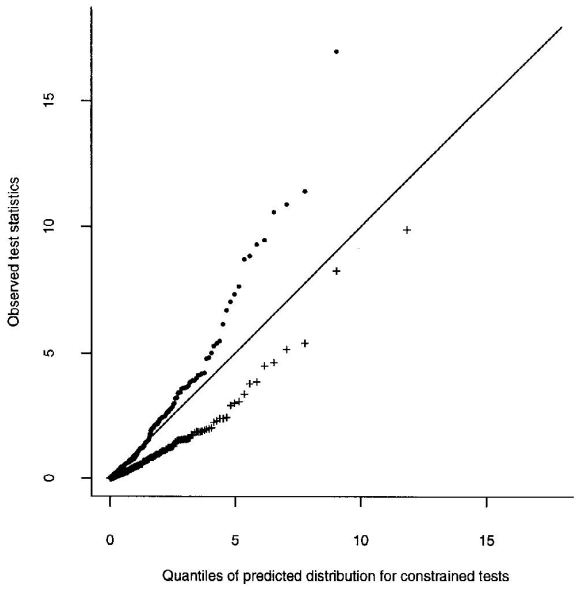

Figure 2 Distribution of weighted and unweighted constrained likelihood-ratio tests for linkage, under the null hypothesis of no linkage: dots denote unweighted tests, and plus signs denote weighted tests. The diagonal line shows the expected distribution corresponding to a mixture of χ^2 with 2 df and 1 df (Holmans 1993), for a fully informative marker.

Table 1

Estimated Design Effects for Three Simulated Data Sets of 1,000 Families, by Number of Affected Siblings per Family

	No. of Affected Siblings in Family			
	Three	Four	Five	Six or More
	Design Effect (No. of Families)			
Fully informative marker:[a]				
Simulation 1	1.043 (248)	1.002 (78)	.996 (37)	1.013 (29)
Simulation 2	1.002 (232)	1.032 (93)	1.005 (39)	.999 (27)
Simulation 3	1.000 (224)	1.004 (88)	1.007 (36)	1.001 (32)
Marker with four equally frequent alleles:[a,b]				
Simulation 1	1.028 (248)	1.062 (78)	1.428 (37)	1.436 (29)
Simulation 2	1.050 (232)	1.294 (93)	1.172 (39)	1.225 (27)
Simulation 3	1.075 (224)	1.207 (88)	1.352 (36)	1.846 (32)
	Inverse of Conventional Weights, $1/w_k$			
$k/2$	1.5	2.0	2.5	3.0
Hodge (1984)[c]	1.214	1.553	1.912	2.278

[a] The three simulations are generated under the same model with different random number seeds and show sampling variability.

[b] Parental-marker typing is unknown, and identity by descent is estimated on the basis of allele frequencies.

[c] Inverse of the weightingt that Hodge (1984) suggested on the basis of the Shannon selective information.

fall inside the plausible triangle. The conventionally weighted tests, again, follow a biased distribution and will be very conservative. The unweighted statistics have a distribution close to the expected distribution, except in the upper tail. As developed here, the estimated design effects will be exactly the same for constrained or unconstrained tests, since they depend not on the estimation method but only on the score elements. However, skewness of the IBD distribution, together with incompletely informative markers, will increase the probability, θ, that the estimated allele sharing will fall inside the plausible triangle, leading to anticonservative tests for an unweighted analysis.

In conclusion, in most studies of affected sib pairs, marker informativity will vary from family to family, parents will sometimes be unavailable for typing, and the skewness of the pairwise IBD distribution for large sibships will persistently affect the likelihood ratio–test statistic's distribution even in large samples. Therefore, the distribution of the constrained or unconstrained likelihood-ratio tests may be slightly anticonservative when large families are not down-weighted (Abel and Müller-Myhsok 1998). However, conventionally weighted likelihood-ratio tests and MLS results should be considered to be extremely conservative if there are numerous families with many affected sibs. Further work is needed to develop a weight that can be empirically estimated and that is appropriate for likelihood-ratio tests even with large sibships. Accurate assessment of sib-pair linkage–significance levels, therefore, should be obtained (*a*) by simulation for likelihood-ratio tests, (*b*) from the means or proportions tests that are unbiased under the null hypothesis, or (*c*) from a method that uses the family as a unit rather than as component sib pairs. Of course, when linkage exists, there may be differences between the powers of these different approaches, so the choice of test statistic should be based on additional considerations.

Acknowledgments

During her doctoral studies, C.M.T.G. was supported by the Health Research Personnel Development Program of the Ontario Ministry of Health, and currently she is supported by a postdoctoral fellowship from the Medical Research Council of Canada. S.B.B. is a National Health Research Scholar of the National Health Research Development Program. This work was also supported by a Collaborative Project Grant from the Natural Sciences and Engineering Research Council of Canada. We thank the reviewers for insightful help and comments.

CELIA M. T. GREENWOOD[1] AND SHELLEY B. BULL[2]
[1]*Department of Human Genetics, McGill University, Montreal; and* [2]*Samuel Lunenfeld Research Institute of Mount Sinai Hospital and Department of Public Health Sciences, University of Toronto, Toronto*

References

Abel L, Müller-Myhsok B (1998) Robustness and power of the maximum-likelihood–binomial and maximum-likelihood–score methods, in multipoint linkage analysis of affected-sibship data. Am J Hum Genet 63:638–647

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

Daly MJ, Lander ES (1996) The importance of being independent: sib pair analysis in diabetes. Nat Genet 14:131–132

Davis S, Weeks DE (1997) Comparison of nonparametric statistics for detection of linkage in nuclear families: single-marker evaluation. Am J Hum Genet 61:1431–1444

Eliasziw M, Donner A (1991) A generalized non-iterative approach to the analysis of family data. Ann Hum Genet 55: 77–90

Greenwood CMT (1998) Models for variable and age-dependent penetrances in genetic linkage analysis. PhD thesis, Graduate Department of Community Health, University of Toronto, Toronto

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

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

Holmans P, Clayton D (1995) Efficiency of typing unaffected relatives in an affected-sib-pair linkage study with single-locus and multiple tightly linked markers. Am J Hum Genet 57:1221–1232

Kong A, Frigge M, Bell GI, Lander ES, Daly MJ, Cox NJ (1997) Diabetes, dependence, asymptotics, selection and significance. Nat Genet 17:148

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

Meunier F, Philippi A, Martinez M, Demenais F (1997) Affected sib-pair tests for linkage: type I errors with dependent sib-pairs. Genet Epidemiol 14:1107–1111

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

Scott AJ, Holt D (1982) The effect of two-stage sampling on ordinary least squares methods. J Am Stat Assoc 77: 848–854

Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Sham PC, Zhao JH, Curtis D (1997) Optimal weighting scheme for affected sib-pair analysis of sibship data. Ann Hum Genet 61:61–69

Suarez BK, Hodge SE (1979) A simple method to detect linkage for rare recessive diseases: an application to juvenile diabetes. Clin Genet 15:126–136

Suarez BK, Van Eerdewegh P (1984) A comparison of three affected-sib-pair scoring methods to detect HLA-linked disease susceptibility genes. Am J Med Genet 18:135–146

Address for correspondence and reprints: Dr. Celia Greenwood, Division of Medical Genetics, Montreal General Hospital, L10 -109, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada. E-mail: celia@utstat.utoronto.ca

0002-9297/99/6404-0045$02.00