Am. J. Hum. Genet. 66:1661–1668, 2000

Power Comparison of Parametric and Nonparametric Linkage Tests in Small Pedigrees

Pak Chung Sham,[1,2] Ming-Wei Lin,[1] Jing Hua Zhao,[1] and David Curtis[3]

Departments of [1]Psychiatry and [2]Biostatistics and Computing, Institute of Psychiatry, De Crespigny Park, Denmark Hill, and [3]Department of Adult Psychiatry, The Royal London Hospital, Whitechapel, London

When the mode of inheritance of a disease is unknown, the LOD-score method of linkage analysis must take into account uncertainties in model parameters. We have previously proposed a parametric linkage test called "MFLOD," which does not require specification of disease model parameters. In the present study, we introduce two new model-free parametric linkage tests, known as "MLOD" and "MALOD." These tests are defined, respectively, as the LOD score and the admixture LOD score, maximized (subject to the same constraints as MFLOD) over disease-model parameters. We compared the power of these three parametric linkage tests and that of two nonparametric linkage tests, NPL_{all} and NPL_{pairs}, which are implemented in GENEHUNTER. With the use of small pedigrees and a fully informative marker, we found the powers of MLOD, NPL_{all}, and NPL_{pairs} to be almost equivalent to each other and not far below that of a LOD-score analysis performed under the assumption the correct genetic parameters. Thus, linkage analysis is not much hindered by uncertain mode of inheritance. The results also suggest that both parametric and nonparametric methods are suitable for linkage analysis of complex disorders in small pedigrees. However, whether these results apply to large pedigrees remains to be answered.

Introduction

The traditional LOD-score method of linkage analysis is designed for the detection of a disease locus characterized by a disease-gene frequency and three penetrance parameters. The method is robust to minor misspecification of disease-model parameters, is powerful for detection of genes of major effect, and is applicable to pedigrees of variable structure (Clerget-Darpoux 1986). However, it is not directly applicable when the disease model is unknown, as is the case for many common heritable disorders. Several methods of "model-free" linkage analysis have been proposed for these genetically complex traits.

There are two broad classes of model-free methods of linkage analysis. First, there are modifications of the classic LOD-score method, in which gene frequencies and penetrances are treated as nuisance parameters. The common practice of conducting several LOD-score analyses under a range of genetic models and of then applying a simple adjustment to the largest LOD score is an example of this approach (Hodge et al. 1997; Greenberg et al. 1998). An extension of this method involves formally maximizing the LOD score over both the recombination fraction and the disease-model parameters, to obtain the so-called "MOD score" (Risch 1984; Greenberg 1989; Hodge and Elston 1994; Rice et al. 1995). The MOD score can be generalized to account for locus heterogeneity, by means of maximizing over the proportion of families with linkage in addition to the other disease-model parameters.

One of the problems with the MOD score is that its sampling distribution under the null hypothesis is uncertain. This problem motivated the development of the model-free LOD (MFLOD), in which the numerator and denominator of the likelihood ratio are separately maximized so that the resulting statistic has a regular χ^2 distribution (Curtis and Sham 1995). The disease-model parameters are constrained to give the population morbid risk as a rough adjustment for ascertainment. A limited range of fully dominant and recessive models is considered in order to reduce the computational burden. When applied to a single chromosomal location, the MFLOD was shown to be a valid χ^2 test with 1 df.

Although the MFLOD appears to give good power for the detection of linkage in many circumstances, we have noticed that, occasionally, its value is far less than that of the admixture LOD score maximized over disease-model parameters, when the same constraints on disease-model parameters are applied. Interestingly, subjecting the disease-model parameters to these constraints has the desirable consequence that the resulting

Received August 22, 1997; accepted for publication February 8, 2000; electronically published April 11, 2000.

Address for correspondence and reprints: Dr. Pak Chung Sham, Department of Psychiatry, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom. E-mail: p.sham@iop.kcl.ac.uk

0002-9297/2000/6605-0020$02.00

maximized LOD score or maximized admixture LOD score is approximately proportional to χ^2 random variables, as we show in the present study. We use the notations "MLOD" and "MALOD" to denote the maximized LOD score and the maximized admixture LOD score, over disease-model parameters subject to constraints that are the same as those for MFLOD. Since their asymptotic distributions are approximately proportional to χ^2 random variables with few degrees of freedom, we consider MLOD and MALOD to be potentially attractive parametric tests of linkage when the disease model is uncertain.

The second approach to model-free linkage analysis is based on excessive sharing of marker alleles among family members that are concordant for the disease phenotype. This nonparametric approach was initially developed as a method for analysis of affected sib pairs (Suarez et al. 1978), and it has recently been extended for use in pedigrees (Weeks and Lange 1988; Sandkujil 1989; Holmans 1993; Curtis and Sham 1994; Davies et al. 1996). The latest of these methods are based on test statistics known as "NPL_{all}" (which measures excessive allele sharing among a set of affected pedigree members) and "NPL_{pairs}" (which measures excessive allele sharing between pairs of affected relatives), and they have been implemented in the GENEHUNTER program (Kruglyak et al. 1996). The results of preliminary power studies suggest that the NPL_{all} and NPL_{pairs} statistics are, in some situations, almost as powerful as the LOD score calculated under the true disease model (Kruglyak et al. 1996).

There has been little systematic evaluation of the performance of parametric and nonparametric linkage tests under a range of conditions, except in affected sib pairs (Hodge 1998). In the present study, we attempt to compare the power of parametric and nonparametric linkage methods that are applicable to general pedigree and multipoint data, under a range of genetic models and pedigree structures.

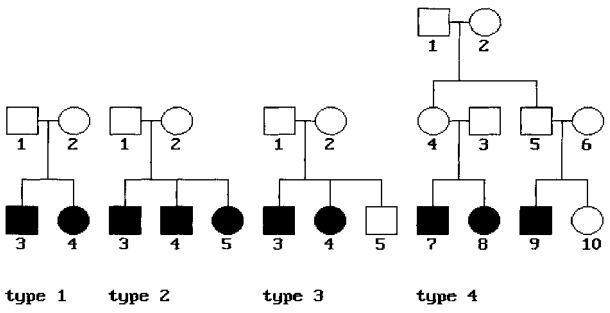

Figure 1 Four different types of pedigree structure

Table 1

The Four Genetic Models Used in Power Analysis

Model	f_2	f_1	f_0	q	K	c_2	c_1	c_0
CR	.50	.005	.005	.100	.01	.503	.090	.407
CD	.50	.500	.005	.005	.01	.001	.501	.498
MG1	.80	.200	.050	.130	.10	.140	.468	.392
MG2	.45	.150	.050	.207	.10	.193	.493	.315

NOTE.—f_2, f_1, and f_0 are the probabilities of disease, given 2, 1, and 0 copies of the disease alleles, respectively; q is the disease-allele frequency; K is the population morbid risk; and c_2, c_1, and c_0 are the probabilities, among affected individuals, of having 2, 1, or 0 copies of the disease alleles.

Pedigrees and Methods

Pedigree Types

We considered four types of pedigree structures (fig. 1). Pedigree type 1 includes two affected siblings with parents of unknown phenotype. Pedigree type 2 includes three affected siblings with parents of unknown phenotype. Pedigree type 3 includes two affected siblings and one unaffected sibling with parents of unknown phenotype. Pedigree type 4 includes three generations in which all family members are of unknown phenotype, except for the youngest generation, which consists of two sibships (who are first cousins to each other), one of which has two affected members and the other of which has one affected member and one unaffected member. These four types of pedigrees were chosen to investigate factors that may influence the relative performance of the tests, such as the number of affected and unaffected individuals and the presence or absence of distantly related affected individuals.

Genetic Models

The four genetic models considered in the present study were the common recessive (CR), common dominant (CD), modest-genetic-effect (MG1), and minor-genetic-effect (MG2) models. The values of the parameters of these models are shown in table 1. These parameters are such that, for the CR and CD models, the morbid risk in the population (K) is .01, and, for the MG1 and MG2 models, $K = 0.1$. Each model was considered under two situations: locus homogeneity (with linkage between disease and markers in all families) and locus heterogeneity (with linkage between disease and markers in 50% of families).

Genetic Markers

Since the aim of the present study was to compare test statistics, we considered the simplest and most-favorable case of recombination fraction (θ) 0 between a completely informative marker locus and the disease locus. In practice, this meant that all the alleles among

the founders of a pedigree were distinguishable from each other. Marker genotypes were assumed to be available for every member of the pedigree, so that allele frequencies were irrelevant in the calculation of likelihoods or nonparametric statistics.

Test Statistics

The test statistics compared are as follows: LOD (the LOD score calculated under the true disease model), ALOD (the admixture LOD score calculated under the true disease model), MLOD, MALOD, MFLOD (Curtis and Sham 1995), NPL_{all}, and NPL_{pairs}. LOD and ALOD, which are calculated under the true disease model, are included for comparison with the other, model-free tests. In the calculation of MLOD, MALOD, and MFLOD, the disease-model parameters are constrained to give the specified K and to be either fully dominant or fully recessive. These constraints imply that, of the parameters q (disease-gene frequency), f_0, f_1, and f_2 (penetrances), only f_1 remains free (Curtis and Sham 1995). For a penetrance parameter f_1 and a fixed K, the other penetrance parameters are given by $f_0 = f_1$ and $f_2 = 1 - f_1(1 - K)/K$, when $f_1 \leq K$, and by $f_2 = f_1$ and $f_0 = (1 - f_1)K/(1 - K)$, when $f_1 > K$; the disease-gene frequency q is given by $q^2 f_2 + 2q(1 - q)f_1 + (1 - q)^2 f_0 = K$. Therefore, the only free parameter associated with MLOD is f_1, which is allowed to vary between 0 and 1, with a null-hypothesis value of $f_1 = K$. (2ln10)MLOD is therefore asymptotically χ^2 with 1 df (see Appendix). Similar constraints are applied to the MALOD statistic, which is then characterized by two free parameters, the penetrance f_1 and the admixture proportion α. These two parameters are completely confounded under the null hypothesis (which can be specified by either $f_1 = K$ or $\alpha = 0$); however, considering (2ln10)MALOD to be χ^2 with 2 df should yield a conservative test. The (2ln10)MFLOD statistic is asymptotically χ^2 with 1 df, since the only parameter that is free in the numerator likelihood but is fixed (to 0) in the denominator is likelihood α (Curtis and Sham 1995). MFLOD also differs from MLOD and MALOD in that it is the logarithm of a ratio of the joint likelihoods of marker and disease phenotypes, rather than the logarithm of a ratio of the conditional likelihoods of marker phenotypes, given disease phenotypes. The definitions of NPL_{all} and NPL_{pairs} are given elsewhere (Kruglyak et al. 1996).

Power Calculations

Power calculations were performed for each test statistic and each pedigree type. All possible marker-genotype configurations of each family type were enumerated. For example, for an affected sib pair, each sibling has four possible marker genotypes, so that there are 16 possible marker-genotype configurations for the sib pair as a whole. Each configuration was subjected to likelihood calculations, by use of VITESSE (O'Connell and Weeks 1995), and nonparametric calculations, by use of GENEHUNTER (Kruglyak et al. 1996). Likelihoods were obtained for all possible configurations, over a fine grid of values for f_1, under the null hypothesis ($\theta = .5$) and again under the alternative hypothesis ($\theta = 0$), for each of the four genetic models. These likelihoods provide the contributions of the different marker-genotype configurations to the expected values of the various parametric statistics. Under the null hypothesis, all possible configurations occur with equal probabilities. Under the alternative hypothesis of linkage in all families, the configurations occur with probabilities proportional to their likelihoods under the assumption of complete linkage. Under the alternative hypothesis of linkage in 50% of families, the configurations occur with probabilities that are 50:50 averages between equal probabilities and those calculated under the assumption that there is linkage in all families. These probabilities were used to calculate the expected values of the log likelihoods, the LOD scores, and the ALOD scores, over a fine grid of values for f_1, under the null hypothesis and the two alternative hypotheses (linkage in 100% and in 50% of families). The maximum expected LOD score over f_1 is the noncentrality parameter (per pedigree) of MLOD, and the maximum expected admixture LOD score over f_1 and α is the noncentrality parameter (per pedigree) of MALOD. The difference in the maximum expected log likelihood over f_1 and α and the maximum expected log likelihood over f_1 (with α set at 0) is the noncentrality parameter (per pedigree) of the MFLOD. For MLOD and MFLOD, which have 1 df, the required sample size for 90% power, at $P = .0001$, is 26.76 divided by the noncentrality parameter per pedigree. For MALOD, which may be regarded as having 2 df, the required sample size for 90% power, at $P = .0001$, is 29.92 divided by the noncentrality parameter per pedigree.

The means and standard deviations of the NPL_{all}, NPL_{pairs}, LOD, and ALOD, over all possible marker-genotype configurations, were calculated for the null hypothesis and for the two alternative hypotheses. These statistics are asymptotically normal, so that the required sample size for a certain power and a certain significance level can be expressed as a function of the means and standard deviations under the null and alternative hypotheses. When the test statistic is scaled to have a mean of 0 and variance of 1 under the null hypothesis, then the required sample size for 90% power and .0001 significance, when the mean and standard deviation under the alternative hypothesis are m and v, is given by $n = [(3.719 + 1.282\sqrt{v})/m]^2$.

We also performed power calculations for samples consisting of a mixture of the four pedigree types. The proportions of pedigree types 1, 2, 3 and 4 in a sample were assumed to be .5, .2, .2, and .1, respectively. The expected log-likelihood function of a mixture of pedigree types is simply a weighted average of the log-likelihood functions of the constituent pedigree types. Noncentrality parameters per pedigree and required sample sizes for 90% power and .0001 significance, for MLOD, MALOD, and MFLOD, were calculated from expected log-likelihood functions.

Required sample sizes of the NPL_{all}, NPL_{pairs}, LOD, and ALOD statistics, for a mixture of pedigree types, were also calculated from their means and variances under the null and alternative hypotheses. The mean and variance of a mixture of four distributions are $m = \Sigma_{i=1}^{4} p_i m_i$ and $v = \Sigma_{i=1}^{4} p_i (v_i + m_i)^2 - m^2$, where p_i, m_i, and v_i are the mixing proportion, mean, and variance of the ith distribution.

Validation of Asymptotic Sampling Distributions

Asymptotic significance and power depend on the accuracy of the χ^2 and normal approximations. To check the validity of the tests, on the basis of the large-sample theory, we generated random samples of pedigrees of various possible marker-genotype configurations, from the probability distribution of these configurations, under the null hypothesis. The numbers of pedigrees in each sample are such that, according to asymptotic theory, a LOD-score test should have 90% power at $P = .0001$. The different test statistics were calculated for 10,000 simulated samples, to provide empirical sampling distributions of the test statistics. These empirical distributions were then compared with the corresponding theoretical χ^2 or normal distributions, to see whether the asymptotic χ^2 or normal tests are accurate or whether they are conservative or liberal.

Results

Table 2 shows the estimates of the numbers of pedigrees required by the LOD, MLOD, MALOD, MFLOD, NPL_{all}, and NPL_{pairs} tests for 90% power (at $P = .0001$) under the four disease models (under the assumption of locus homogeneity) for the four types of pedigrees (and a mixture of families). As expected, LOD is the most-efficient test. However, it is notable that the required sample sizes of the model-free tests are, in general, not much larger than those of LOD. Among the model-free tests, NPL_{all}, NPL_{pairs}, and MLOD have slightly better power than do MALOD and MFLOD. The patterns of results, with regard to genetic model and pedigree structure, are as expected, with greater genetic-effect size and larger pedigrees being more favorable for the detection of linkage than are smaller genetic-effect size and smaller pedigrees. For MG1 and MG2, pedigree type 2 (sibships with three affected members) was the most efficient of the four pedigree types.

Table 2

Estimated Required Sample Sizes for 90% Power, at $P = .0001$, When There Is Complete Linkage to a Fully Informative Marker in All Families

Model and Family[a]	Estimated Required Sample Sizes, for 90% Power, for					
	LOD	NPL_{pair}	NPL_{all}	MLOD	MALOD	MFLOD
CR:						
1	18	20	20	22	24	32
2	10	10	10	13	15	17
3	16	19	19	20	22	42
4	13	14	14	15	17	48
Mixed	15	16	17	18	20	34
CD:						
1	50	52	52	51	57	51
2	16	18	18	20	22	20
3	44	52	52	45	50	72
4	7	9	7	10	11	13
Mixed	28	32	31	29	32	39
MG1:						
1	285	285	285	312	348	311
2	80	80	80	91	102	94
3	245	312	312	275	307	352
4	124	131	132	145	162	280
Mixed	173	192	192	195	218	264
MG2:						
1	626	626	626	677	755	675
2	199	199	199	224	250	224
3	582	641	641	636	710	1,022
4	300	309	315	342	382	771
Mixed	405	437	438	449	502	666

[a] "Mixed" refers to a mixture of pedigree types 1, 2, 3, and 4 in the proportions 50%, 20%, 20%, and 10%.

Under an admixture model in which 50% of families show linkage, a similar pattern of results emerges (table 3). Again, ALOD is the most-powerful test and is closely followed by NPL_{all} and NPL_{pairs} and MLOD. It is perhaps surprising that MALOD does not perform better than MLOD. This may be due to the use of a conservative null distribution (χ^2 distribution with 2 df) for MALOD.

None of the test statistics was drastically conservative or liberal (table 4). As expected, referring (2ln10) MALOD to a χ^2 distribution with 2 df leads to a conservative test. The large-sample-theory tests for LOD, ALOD, NPL_{all}, and NPL_{pairs} tend to be somewhat liberal, especially for pedigrees with multiple affected members and for moderately large genetic effects. In contrast, the asymptotic sampling distribution for MLOD appears to be neither conservative nor liberal.

Table 3

Estimated Required Sample Sizes for 90% Power, at $P = .0001$, When There is Complete Linkage to a Fully Informative Marker in 50% of Families

Model and Family[a]	Estimated Required Sample Sizes, for 90% Power, for					
	ALOD	NPL_{pair}	NPL_{all}	MLOD	MALOD	MFLOD
CR:						
1	74	92	92	84	94	119
2	35	45	45	44	48	63
3	65	87	87	77	84	172
4	44	57	57	56	58	157
Mixed	57	73	73	72	74	131
CD:						
1	227	228	228	241	270	241
2	79	80	80	88	99	88
3	191	228	228	224	228	284
4	26	40	33	35	37	47
Mixed	111	138	133	133	134	149
MG1:						
1	1,139	1,151	1,151	1,245	1,390	1,243
2	305	315	315	345	379	357
3	965	1,258	1,258	1,105	1,206	1,376
4	468	512	509	538	599	1,008
Mixed	663	764	763	755	841	1,041
MG2:						
1	2,503	2,515	2,515	2,701	3,018	2,703
2	771	781	781	864	960	860
3	2,317	2,577	2,577	2,514	2,806	4,036
4	1,159	1,212	1,224	1,288	1,439	2,887
Mixed	1,585	1,739	1,740	1,761	1,964	2,675

[a] "Mixed" refers to a mixture of pedigree types 1, 2, 3, and 4 in the proportions 50%, 20%, 20%, and 10%.

Discussion

The results of the present study show that uncertain mode of inheritance is not a serious problem for linkage analysis. In small pedigrees—such as those considered here—both nonparametric (NPL_{all} and NPL_{pairs}) and parametric (MLOD) tests performed almost as well as did the LOD or ALOD statistics, for which the true genetic model was assumed. If a nonparametric test is to be used for a complex trait, however, then the conceptually and computationally simpler NPL_{pairs} test is almost as powerful as the more-complicated NPL_{all}, and it may be less liberal in large pedigrees.

The poor performance of MFLOD, relative to MLOD and MALOD, can be explained by the observation that parameter estimates in the numerator of the likelihood ratio in MFLOD are often quite far from the true values. It appears that constraining the disease-model parameters to be compatible with the morbid risk in the population is not sufficient to produce realistic parameter estimates for a minor locus. The estimate of f_2 is often much larger than the true value. As a result, the test statistic is obtained at the wrong region of the parameter space, with consequent loss of power.

Use of the MLOD and MALOD, since they are ratios of conditional likelihoods, enables one to obtain unbiased estimates of the disease-model parameters. This enables the statistics to be obtained from the correct region of the parameter space, even for a gene of minor effect. As expected, MLOD is superior to MALOD in the absence of locus heterogeneity. More surprisingly, even under substantial locus heterogeneity, MLOD remains superior to MALOD. The reason for this is that, since the penetrance and admixture are strongly confounded in small pedigrees, the extra degree of freedom in MALOD may be redundant. Nevertheless, the overall pattern of results favors the choice of MLOD as the single best model-free parametric linkage test for small pedigrees.

The results of the present study confirm that both NPL statistics provide powerful tests for linkage when marker information content is complete. Recently, Kong and Cox (1997) have shown that the original NPL statistics are conservative when marker information is incomplete, especially at positions between markers, resulting in loss of power. They have proposed modified NPL statistics that are not conservative and that, therefore, are more powerful for low levels of marker information content and positions between markers. When marker information content is complete (as is assumed in the present study), the modified NPL statistics have power equal to that of the original NPL statistics. The modified NPL tests, which have been implemented in GENEHUNTER-PLUS, are based on a statistical model with a single parameter. It has been shown that the nonparametric mean test for affected sib pairs is equivalent to parametric LOD-score analysis under a recessive model (Knapp et al. 1994). The development of model-based nonparametric tests in GENEHUNTER-PLUS is an important development that brings parametric and nonparametric methods closer together. Further work is necessary to examine the advantages and disadvantages of alternative specifications and parameterizations of the statistical model.

In the present study, we used a single, infinitely polymorphic marker to approximate a fully informative multipoint analysis. We should point out that current nonparametric methods, such as the NPL statistics in GENEHUNTER or the modified NPL statistics in GENEHUNTER-PLUS, are inherently suited to multipoint analysis, since these statistics assess, at a test position, the evidence for distortion in allele sharing among affected relative pairs. Standard two-point parametric methods use the recombination fraction between disease and marker as the parameter, whereas the disease model is prespecified. The generalization of this

Table 4

Proportion of Replicate Samples (of 10,000) Simulated under the Null Hypothesis (Significant at $P < .001$)

Model and Family[a]	Proportion of Replicate Samples for						
	LOD	ALOD	NPL_{pair}	NPL_{all}	MLOD	MALOD	MFLOD
CR:							
1	21	13	7	7	18	3	14
2	5	26	41	41	10	2	6
3	20	14	11	11	14	6	7
4	16	35	21	25	10	3	11
CD:							
1	4	8	10	10	8	1	7
2	7	9	1	11	5	0	3
3	5	10	10	10	10	3	5
4	37	43	35	57	15	9	4
MG1:							
1	5	6	5	5	2	0	1
2	22	14	22	22	8	1	6
3	8	12	7	7	5	2	5
4	12	12	12	17	9	1	7
MG2:							
1	6	15	6	6	6	1	5
2	9	9	9	9	3	1	3
3	8	13	12	12	5	5	5
4	5	20	9	9	6	2	3

[a] "Mixed" refers to a mixture of pedigree types 1, 2, 3, and 4 in the proportions 50%, 20%, 20%, and 10%.

procedure to multipoint analysis is problematic, since the recombination fraction ceases to be an effective parameter (Risch and Giuffra 1992). In GENEHUNTER, the proportion of linked pedigrees provides an alternative parameter with which to assess the evidence for a disease gene at a test location. The MLOD statistic described in the present study would also provide an alternative one-parameter test with which to assess the evidence for a disease gene in a multipoint analysis. The adoption of this statistic for multipoint parametric linkage analysis would obviate the need for two-point parametric linkage analysis, except to reduce computational burden in the initial stages of a genome scan.

A limitation of this study is that all the calculations were based on a generalized single-locus model rather than on an oligogenic or mixed model. Since complex disorders are likely to be determined by multiple loci, the single-locus model is biologically unrealistic. However, the generalized single-locus model is probably an adequate approximation of an oligogenic model, when one is trying to map one disease gene at a time, rather than when one is trying to map two or more disease genes simultaneously (Greenberg and Hodge 1989; Greenberg 1990; MacLean et al. 1993). Another limitation is that all our analyses have been performed on small pedigrees. The behaviors of the different tests in large pedigrees remain to be explored by means of analytic or simulation studies.

Acknowledgments

This work was supported by Wellcome Trust grant 055379, Medical Research Council grant G9700821, and National Institutes of Health grant EY-12562. M.-W.L. was supported by the Ministry of Education, Taiwan, Republic of China. We are grateful to an anonymous referee, for valuable comments that resulted in substantial revision of the Discussion section.

Appendix

Asymptotic Distributions of MLOD and MALOD under the Null Hypothesis

The MLOD and MALOD at a fixed map position, given a set of pedigree data, are defined as the maximum LOD score and the maximum admixture LOD score, respectively, at that position, over a restricted set of transmissions models. The restriction of transmission models (described in Curtis and Sham 1995) involves, first, the specification of K. A single major locus effect is parameterized as the gene frequency and penetrance vector q and

$\mathbf{F} = (f_0,f_1,f_2)$, respectively. For each value of **F**, there is a unique value for q that yields the correct K. The set of models is further limited to those in the straight lines lying between (0,0,1) (Mendelian recessive) and (K,K,K) (null effect) and between (K,K,K) and (0,1,1) (Mendelian dominant). All parameters p and **F** of the transmission model can therefore be specified as functions of f_1, which takes values between 0 and 1. For a given test position of the disease locus, the log likelihood of the data set is a function of f_1 and α (the proportion of affected pedigrees in which that locus exerts an effect). The test position is denoted as t, with $\theta = t$ implying that the disease locus is placed at the test position, relative to the marker or markers, and with $\theta = .5$ implying that it is unlinked to any of the markers.

MLOD is the logarithm of the maximum over f_1 (from 0 to 1) of the quantity:

$$\begin{aligned} LR &= P(D,\ M;\ \theta = t,\ f_1)/P(D,\ M;\ \theta = .5,\ f_1) \\ &= P(M \mid D;\ \theta = t,\ f_1)/P(M) \\ &= P(M \mid D;\ \theta = t,\ f_1)/P(M \mid D;\ \theta = t,\ f_1 = K)\ , \end{aligned}$$

which is the ratio of the conditional probabilities of observing the marker data, given the disease data under the hypotheses of some genetic effect ($f_1 \neq K$) and null effect ($f_1 = K$). Since these hypotheses are nested, the asymptotic distribution of $2\ln(LR)$ (i.e., MLOD*2ln10) is approximately χ^2_1 (two-tailed, since f_1 can be greater than or less than K).

MALOD is the logarithm of the maximum over f_1 (from 0 to 1) of the quantity:

$$\begin{aligned} LR &= P(D,\ M;\ \theta = t,\ \alpha,\ f_1)/P(D,\ M;\ \theta = 0.5,\ \text{or}\ \alpha = 0,\ f_1) \\ &= P(M \mid D;\ \theta = t,\ \alpha,\ f_1)/P(M) \\ &= P(M \mid D;\ \theta = t,\ \alpha,\ f_1)/P(M \mid D;\ \theta = t,\ \alpha = 0,\ \text{or}\ f_1 = K) \end{aligned}$$

Again, LR is the ratio of the conditional probabilities of observing the marker data, given the disease data, under linkage and nonlinkage. The numerator is maximized over f_1 and α, with the null hypothesis represented by $f_1 = K$ and/or by $\alpha = 0$. Because these two parameters are confounded under the null hypothesis, a test comparing $2\ln(LR)$ with a 50:50 mixture of χ^2_2 and χ^2_0 should be somewhat conservative.

References

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986) Effects of misspecifying genetic parameters in LOD score analysis. Biometrics 42:393–399

Curtis D, Sham PC (1994) Using risk calculation to implement an extended relative pair analysis. Ann Hum Genet 58:151–162

Curtis D, Sham PC (1995) Model-free linkage analysis using likelihoods. Am J Hum Genet 57:703–716

Davies S, Schroeder M, Goldin LR, Weeks DE (1996) Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am J Hum Genet 58:867–880

Greenberg DA (1989) Inferring mode of inheritance by comparison of LOD scores. Am J Med Genet 34:480–486

Greenberg DA (1990) Linkage analysis assuming a single-locus mode of inheritance for traits determined by two loci: inferring mode of inheritance and estimating penetrance. Genet Epidemiol 7:467–479

Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analysis. Am J Hum Genet 63:870–879

Greenberg DA, Hodge SE (1989) Linkage analysis under "random" and "genetic" reduced penetrance. Genet Epidemiol 6:259–264

Hodge SE (1998) Exact ELODs and exact power for affected sib pairs analyzed for linkage under simple right and wrong models. Am J Med Genet 81:66–72

Hodge SE, Abreu PC, Greenberg DA (1997) Magnitude of type 1 error when single-locus linkage analysis is maximized over models: a simulation study. Am J Hum Genet 60: 217–227

Hodge SE, Elston RC (1994) LODs, WRODs, and MODs: the interpretation of LOD scores calculated under different models. Genet Epidemiol 11:329–342

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

Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. 2. Relationship between affected sib-pair tests and LOD score analysis. Hum Hered 44:44–51

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

MacLean CJ, Sham PC, Kendler KS (1993) Joint linkage of multiple loci for a complex disorder. Am J Hum Genet 53: 353–366

O'Connell JR, Weeks DE (1995) The VITESSE algorithm for

rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. Nat Genet 11:402–408

Rice JP, Neuman RJ, Hoshaw SL, Daw EW, Gu C (1995) TDT with covariates and genomic screens with MOD scores: their behavior on simulated data. Genet Epidemiol 12:659–664

Risch N (1984) Segregation analysis incorporating linkage markers. I. Single locus models with an application to type I diabetes. Am J Hum Genet 36:363–386

Risch N, Giuffra L (1992) Model misspecification and multipoint linkage analysis. Hum Hered 42:77–92

Sandkuijl LA (1989) Analysis of affected sib-pairs using information from extended families. In: Eston RC, Spencer MA, Hodge SE, MacCluer JW (eds) Multipoint mapping and linkage based upon affected pedigree members: genetic analysis workshop 6. Alan R. Liss, New York

Suarez BK, Rice J, Reich T (1978) The generalized sib pair IBD distribution: its use in the detection of linkage. Ann Hum Genet 42:87–94

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326