

# Conclusions of LOD-Score Analysis for Family Data Generated under Two-Locus Models

Marie-Hélène Dizier, Marie-Claude Babron, and Françoise Clerget-Darpoux

Unité de Recherches d'Epidémiologie Génétique, INSERM U155 Paris

## Summary

The power to detect linkage by the LOD-score method is investigated here for diseases that depend on the effects of two genes. The classical strategy is, first, to detect a major-gene (MG) effect by segregation analysis and, second, to seek for linkage with genetic markers by the LOD-score method using the MG parameters. We already showed that segregation analysis can lead to evidence for a MG effect for many two-locus models, with the estimates of the MG parameters being very different from those of the two genes involved in the disease. We show here that use of these MG parameter estimates in the LOD-score analysis may lead to a failure to detect linkage for some two-locus models. For these models, use of the sib-pair method gives a non-negligible increase of power to detect linkage. The linkage-homogeneity test among subsamples differing for the familial disease distribution provides evidence of parameter misspecification, when the MG parameters are used. Moreover, for most of the models, use of the MG parameters in LOD-score analysis leads to a large bias in estimation of the recombination fraction and sometimes also to a rejection of linkage for the true recombination fraction. A final important point is that a strong evidence of an MG effect, obtained by segregation analysis, does not necessarily imply that linkage will be detected for at least one of the two genes, even with the true parameters and with a close informative marker.

## Introduction

The classical strategy to detect and localize a gene involved in a disease is first to perform segregation analysis, followed by linkage analysis with genetic markers. The purpose of segregation analysis is to detect a major-

gene (MG) effect by testing restrictive versus more general models with the log-likelihood ratio test. Different models have been proposed: the mixed model (Morton and MacLean 1974), which includes an MG effect and a polygenic component, and the general transmission model, which includes a MG effect allowing for non-Mendelian transmission (Elston and Stewart 1971). Inclusion of the free-transmission probabilities in the mixed model has then given rise to the unified model (Lalouel et al. 1983). When there is evidence of an MG effect, segregation analysis also provides the estimates of its parameters. Usually, linkage analyses with genetic markers are then performed by the LOD-score method using these MG parameters. Such a strategy has been found successful for diseases that follow a clear Mendelian single-locus segregation pattern. However, its relevance for complex diseases, such as those depending on the interactive effect of two genes (two-locus model) has never been verified, although it is now widely used in this context.

We already showed (Dizier et al. 1993) that, for data generated under models when the interactive effect of two genes is assumed, segregation analysis under the unified model often provides evidence for an MG effect associated with a polygenic component. In addition, the existence of an MG effect is supported by evidence of transmission and agreement with the hypothesis of Mendelian transmission. Accordingly, there is no means of detecting that the effect of an MG does not correspond to the correct model. Furthermore the MG parameter estimates may not correspond to the characteristics of either of the two genes involved in the disease (Dizier et al. 1993). Use of these parameter estimates in further linkage analysis may then affect the conclusions of the analysis.

Our purpose here is to investigate, for a disease depending on the effect of two genes, the conclusions drawn by linkage analysis using the LOD-score method. We were interested in two aspects: first, what is the power of the LOD-score method to detect linkage for each of the two susceptibility genes with a linked marker, using the true parameters of the underlying model? Will linkage be detected for both genes, for only one, or for none? Second, what is the impact of using

Received May 22, 1995; accepted for publication February 26, 1996.

Address for correspondence and reprints: Dr. Marie-Hélène Dizier, Unité de Recherches d'Epidémiologie Génétique, INSERM U155, Université Paris 7, Case 7041, 2 place Jussieu, 75005 Paris.

© 1996 by The American Society of Human Genetics. All rights reserved.  
0002-9297/96/5806-0027\$02.00

the MG parameters estimated by segregation analysis, since it is already known that misspecification of the parameters in linkage analysis can affect detection of linkage? For this question, we generated data under a set of two-locus models, including, first, information on the familial segregation of the disease to perform segregation analysis and, second, information on the cosegregation of the disease and two markers, each linked to one of the susceptibility genes, to perform linkage analysis using the LOD-score method successively with each marker.

Finally, besides the LOD-score method, we investigated the interest of two other linkage analyses in the case of diseases depending on the interactive effect of two genes. First, we compare the power to detect linkage of the sib-pair method to that of the LOD-score method, using the MG parameters obtained by segregation analysis. Second, we propose to apply the linkage-homogeneity test in subsamples differing by the familial disease distribution, using the MG parameters.

## Models

Consider a disease determined by the effect of two genes, located at two unlinked biallelic loci,  $M$  and  $N$ . Each locus is linked to a fully informative marker, markers  $A$  and  $B$ , respectively, for the loci  $M$  and  $N$ , with a recombination fraction of 5%. The two-locus model is defined by the following parameters: the frequencies of the susceptibility alleles at each locus,  $q_1$  and  $q_2$ , and the penetrance matrix  $F = (f_{ij})$ , where  $f_{ij}$  is the probability that an individual is affected, given his joint genotype  $(m_i, n_j)$  at the loci  $M$  and  $N$ . The marginal penetrance vectors for each locus can be obtained as follows: for locus  $M$ ,  $V1 = (f_{i.})$  with  $f_{i.} = \sum_j f_{ij} \Pr(n_j)$ ; for locus  $N$ ,  $V2 = (f_{.j})$  with  $f_{.j} = \sum_i f_{ij} \Pr(m_i)$ ;  $\Pr(m_i)$  and  $\Pr(n_j)$  are, respectively, the probabilities of genotype  $m_i$  at locus  $M$  and of genotype  $n_j$  at locus  $N$ .

It is not possible here to perform an exhaustive study of two-locus models, because of the large number of parameters involved. We first consider models where both susceptibility alleles are necessary to develop the disease, so that penetrance for an individual without a susceptibility allele is equal to zero. In model A both genes are dominant, in model B one gene is dominant, the other recessive, and in model C both genes are recessive. We also consider models where presence of both susceptibility alleles is not necessary to develop the disease (D, E, F, G). For these models, presence of phenocopies is thus allowed in the marginal penetrance vector of each gene. In model D both genes are dominant, in model E and F one gene is recessive, the other dominant, and in model G both genes are recessive. For all models, the allelic frequency for the two genes could be either

equal or different. Penetrances and allelic frequencies are chosen in order to obtain a heritability  $< 1$  under the polygenic model.

The conclusions drawn by segregation analysis for traits due to the effect of two genes are not the purpose of the present paper, since they have already been presented already (Dizier et al. 1993). The models presented here are models for which segregation analysis led to conclude to an MG effect so that the parameters of this gene can be used in linkage analysis.

## Methods

Under a set of two-locus models, we generated the segregation of the disease in a family sample, and, through the expected maximum likelihood approach, we performed segregation analysis using the computer program POINTER (see Dizier et al. 1993, for more details). We considered here the case of nuclear families with four children selected from an affected child with an ascertainment probability equal to .01. We calculated the log likelihoods expected in a given sample  $N_1$ . For that sample size, we then performed the tests of models usually considered in segregation analysis to detect MG effect and polygenic component and to verify evidence of transmission for the major effect and agreement with Mendelian transmission.

Under the same set of two-locus models, cosegregation of the disease and the two markers  $A$  and  $B$  is generated in a sample of size  $N_2$  of nuclear families with two affected sibs. Selecting families by the presence of two affected children has been shown to increase efficiency to detect linkage for complex diseases. Families with more than two children were not considered, because of excessive computing time. The LOD-score method (Morton 1955) was then applied to these data successively for each marker  $A$  and  $B$ .

The probability of each familial configuration,  $\Pr(i)$ , depending on the marker and disease distribution, was calculated under the generating two-locus model. The expected LOD score of a model can be then calculated as follows:  $E[Z(\theta)] = \sum_i \Pr(i) \times Z_i(\theta)$ , where  $Z_i(\theta)$  is the LOD score of a family in configuration  $i$  and  $\theta$  is the recombination fraction between the disease and the marker under study. In a family sample of size  $N_2$ , the expected LOD score is thus equal to  $N_2 \times E[Z(\theta)]$ . A program allows us to calculate  $E[Z(\theta)]$  and to maximize it ( $\max E[Z(\theta)]$ ) by the Gemini procedure (Lalouel 1979).  $\max E[Z(\theta)]$  is compared to the critical limit of 3 to provide evidence for linkage.  $E[Z(\theta)] < -2$  leads to exclude linkage for the corresponding value of  $\theta$ .

Computation of  $\max E[Z(\theta)]$  has been first performed under the correct phenotype-genotype correspondence. We either used the complete two-locus model param-

ters or the one-locus parameters of the studied gene with the marginal penetrance vector. Since these two parameterizations lead to very similar results, only the first one will be presented here. Second, the MG parameters were used to compute  $\max E[Z(\theta)]$ . They were also used to calculate  $E[Z(\theta)]$  for the true value of  $\theta$ .

## Results

For all the models presented here, segregation analysis gives evidence for an MG effect, associated or not with a polygenic component, for a sample size of 300 families. Evidence of transmission and agreement with Mendelian transmission are also verified. The MG parameters estimated under the mixed model, the allelic frequency  $q$ , and the penetrance vector  $V$  ( $q$  and  $V$  are easily obtained from the prevalence, the dominance parameter  $d$ , and the displacement  $t$  given by POINTER) are given in table 1. Table 1 also displays the  $\chi^2$  obtained when testing the absence of the MG effect. Table 2 gives, for all these models, the maximum LOD-score value for each susceptibility gene, expected in a sample size of 100 families, using, first, the true parameters ( $Z_{TR1}$  and  $Z_{TR2}$ , respectively, for the first and second susceptibility gene) and second, the MG parameters ( $Z_{MG1}$  and  $Z_{MG2}$ , respectively, for the first and second susceptibility gene). Table 2 also gives the recombination fraction for each gene, with its linked marker estimated by use of the MG parameters, and, finally, the LOD scores  $Z_{MG1}(\theta_0)$  and  $Z_{MG2}(\theta_0)$  calculated for the true recombination fraction.

### Power to Detect Linkage by the LOD-Score Method Using the True Parameters

For models where segregation analysis gives evidence for an MG, there may be detection of linkage with a linked marker either for the two susceptibility genes (models B2, C1, C2, F1, F2, and G1), for only one (models A2, B1, E1, E2, and G2), or for none (models A1, D1, and D2).

Detection of linkage between a susceptibility gene and a marker linked to this gene depends on the gene characteristics as well as on the recombination fraction between the gene and the marker and on the informativity of the marker. We are mostly interested in the characteristics of the gene. It is already known that, in the absence of phenocopies, the power to detect linkage decreases as the dominance parameter and the allelic frequency increase. This may, however, no longer be true in the presence of phenocopies, which itself leads to a decrease in the power to detect linkage. In the case of two-locus models, detection of linkage for one of the genes depends on the characteristics of that gene but also on those of the other, since the marginal penetrance vector for each gene is a function of the allele frequency and penetrance

values at the other gene. However, the rules described above may still be applied to each gene, when the marginal penetrance vector of that gene are considered. In light of the recessive gene of model B1 and the dominant one of model A1, both genes having the same allelic frequency and penetrance value, we indeed obtained a greater LOD score for the recessive gene ( $Z_{TR1} = 15.32$ ) than for the dominant gene ( $Z_{TR1} = Z_{TR2} = 2.13$ ). Also, in considering of the first genes of models A1 and A2, both genes are dominant with the same penetrance vector, but with allelic frequency equal to .2 and to .05, respectively, and the corresponding  $Z_{TR1}$  values are 2.13 and 7.03.

A more interesting point is that there are models for which linkage is not detectable but for which strong evidence is revealed by way segregation analysis of an MG effect. It is the case of model D where there is no detection of linkage ( $Z_{TR1} = Z_{TR2} = .67$ ) but very strong evidence of an MG effect ( $\chi^2 = 31.2$ , 3 df;  $P = 10^{-6}$ ). Conversely, there are models for which the power to conclude to an MG effect is lower and for which linkage may be evidenced (model B2). There are even models (results not shown here) for which there is evidence of linkage for the two genes but no evidence by segregation analysis of an MG effect. In conclusion, there is no clear correspondence between the power of detecting by segregation analysis an MG and of detecting linkage.

### Impact on LOD-Score Analysis When Using the MG Parameters

Table 2 shows that, for several two-locus models, there is no, or only small, bias in estimation of the LOD score when the MG parameters are used instead of the true parameters. In fact, the models for which the estimate of the LOD score is biased are particularly those where the two genes have a different mode of transmission. It has been shown elsewhere that the LOD score is very sensitive to a misspecification of the parameter  $d$  but not to the parameters  $q$  and  $t$  (Clerget-Darpoux et al. 1986). We already pointed out (Dizier et al. 1993) that, when both genes have the same parameter  $d$ , the one estimated for the MG by segregation analysis is close to the true value. This can be verified here: misspecification of  $d$  occurs only when both genes have a different mode of transmission (i.e., models B, E, and F). For these models, a bias in estimation of the LOD score is obtained for the gene with a  $d$  value more different from that estimated for the MG. Furthermore, this bias is the greatest for models with also asymmetrical allelic frequencies for the two genes (models B2, E2, and F2). This corresponds to models with one recessive and one dominant gene and a higher allelic frequency for the recessive gene. The MG parameter estimates correspond to a codominant transmission, and the bias in the LOD

Table 1

## Parameters Estimated by Segregation Analysis from Data Generated under Two-Locus Models

Generating Model	Penetrance Matrix	Gene Frequencies	Marginal Penetrances	MG Parameters Estimated by POINTER	Absence of MG Effect Tested in 300 Families ( $\chi^2$ , 3 df)
Model A	$F = \begin{bmatrix} .5 & .5 & 0 \\ .5 & .5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .2, q_2 = .2$ (2) $q_1 = .05, q_2 = .2$	$V_1 = V_2 = (.18, .18, 0)$ $\left\{ \begin{array}{l} V_1 = (.18, .18, 0) \\ V_2 = (.048, .048, 0) \end{array} \right\}$	$q = .083, V = (.50, .38, .004)$ $q = .024, V = (.51, .35, .001)$	28.7 ( $P = .3 \times 10^{-5}$ ) 52.6 ( $P < 10^{-6}$ )
Model B	$F = \begin{bmatrix} .5 & .5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .2, q_2 = .2$ (2) $q_1 = .5, q_2 = .1$	$\left\{ \begin{array}{l} V_1 = (.18, 0, 0) \\ V_2 = (.02, .02, 0) \end{array} \right\}$ $\left\{ \begin{array}{l} V_1 = (.095, 0, 0) \\ V_2 = (.125, .125, 0) \end{array} \right\}$	$q = .11, V = (.46, .008, 0)$ $q = .033, V = (.999, .30, .004)$	25.6 ( $P = .12 \times 10^{-4}$ ) 14.6 ( $P = .0022$ )
Model C	$F = \begin{bmatrix} .5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .2, q_2 = .2$ (2) $q_1 = .3, q_2 = .05$	$V_1 = V_2 = (.02, 0, 0)$ $\left\{ \begin{array}{l} V_1 = (.0013, 0, 0) \\ V_2 = (.045, 0, 0) \end{array} \right\}$	$q = .036, V = (.38, .004, 0)$ $q = .043, V = (.63, .04, 0)$	28.7 ( $P = .3 \times 10^{-5}$ ) 20.5 ( $P = .00013$ )
Model D	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & .5 & .5 \\ .5 & .5 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .1$ (2) $q_1 = .2, q_2 = .05$	$V_1 = V_2 = (.5, .5, .095)$ $\left\{ \begin{array}{l} V_1 = (.5, .5, .048) \\ V_2 = (.5, .5, .18) \end{array} \right\}$	$q = .192, V = (.495, .495, 0)$ $q = .242, V = (.496, .496, 0)$	31.2 ( $P = .1 \times 10^{-5}$ ) 22.5 ( $P = .5 \times 10^{-4}$ )
Model E	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & .5 & 0 \\ .5 & .5 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .1$ (2) $q_1 = .2, q_2 = .01$	$\left\{ \begin{array}{l} V_1 = (.5, .095, .095) \\ V_2 = (.5, .5, .005) \end{array} \right\}$ $\left\{ \begin{array}{l} V_1 = (.5, .01, .01) \\ V_2 = (.5, .5, .02) \end{array} \right\}$	$q = .104, V = (.495, .495, .002)$ $q = .037, V = (.999, .495, .003)$	49.6 ( $P < 10^{-6}$ ) 24.4 ( $P = .2 \times 10^{-4}$ )
Model F	$F = \begin{bmatrix} .5 & .5 & 0 \\ .1 & 0 & 0 \\ .1 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .1$ (2) $q_1 = .5, q_2 = .05$	$\left\{ \begin{array}{l} V_1 = (.095, .001, .001) \\ V_2 = (.104, .005, 0) \end{array} \right\}$ $\left\{ \begin{array}{l} V_1 = (.05, 0, 0) \\ V_2 = (.2, .125, 0) \end{array} \right\}$	$q = .059, V = (.334, .007, 0)$ $q = .015, V = (.934, .329, .003)$	18.2 ( $P = .4 \times 10^{-3}$ ) 11.2 ( $P = .011$ )
Model G	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & 0 & 0 \\ .5 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .1$ (2) $q_1 = .05, q_2 = .2$	$V_1 = V_2 = (.5, .005, .005)$ $\left\{ \begin{array}{l} V_1 = (.5, .02, .02) \\ V_2 = (.5, .0013, .0013) \end{array} \right\}$	$q = .135, V = (.506, .001, .001)$ $q = .204, V = (.5, .0004, .0004)$	65.7 ( $P < 10^{-6}$ ) 40.1 ( $P < 10^{-6}$ )

**Table 2**

Results of LOD-Score Analysis on 100 Families Generated under Two-Locus Models

GENERATING MODEL	PENETRANCE MATRIX	GENE FREQUENCIES	RESULTS OF THE LOD-SCORE ANALYSIS USING			
			Two-Locus $Z_{TR}$	MG Parameters		
				$Z_{MG}$	$\theta$	$Z_{MG}(\theta_0)$
Model A	$F = \begin{bmatrix} .5 & .5 & 0 \\ .5 & .5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .2, q_2 = .2$	2.13	2.12	.14	<b>1.44</b>
		(2) $q_1 = .05, q_2 = .2 \begin{cases} (1)^a \\ (2) \end{cases}$	7.03	7.03	.07	<b>6.93</b>
			2.02	1.97	.18	-1.07
Model B	$F = \begin{bmatrix} .5 & .5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .2, q_2 = .2 \begin{cases} (1) \\ (2) \end{cases}$	15.32	15.31	.07	15.08
			2.02	1.35	.26	-11.35
		(2) $q_1 = .5, q_2 = .1 \begin{cases} (1) \\ (2) \end{cases}$	4.05	2.16	.16	.57
Model C	$F = \begin{bmatrix} .5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .2, q_2 = .2$	14.26	14.24	.10	12.04
		(2) $q_1 = .3, q_2 = .05 \begin{cases} (1) \\ (2) \end{cases}$	8.75	8.69	.10	7.60
			26.82	26.27	0	24.37
Model D	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & .5 & .5 \\ .5 & .5 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .2$	.67	.68	.18	-.02
		(2) $q_1 = .2, q_2 = .05 \begin{cases} (1) \\ (2) \end{cases}$	1.21	1.21	.09	1.14
			.06	.06	.30	-1.15
Model E	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & .5 & 0 \\ .5 & .5 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .1 \begin{cases} (1) \\ (2) \end{cases}$	.02	0	.42	-5.32
			4.79	4.79	.06	4.79
		(2) $q_1 = .2, q_2 = .01 \begin{cases} (1) \\ (2) \end{cases}$	5.5	2.63	.15	1.10
Model F	$F = \begin{bmatrix} .5 & .5 & 0 \\ .1 & 0 & 0 \\ .1 & 0 & 0 \end{bmatrix}$	(1) $q_1 = .1, q_2 = .1 \begin{cases} (1) \\ (2) \end{cases}$	10.03	9.64	.13	6.67
			6.39	5.95	.17	-.11
		(2) $q_1 = .5, q_2 = .5 \begin{cases} (1) \\ (2) \end{cases}$	3.72	1.71	.19	-1.77
Model G	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & 0 & 0 \\ .5 & 0 & 0 \end{bmatrix}$		6.98	6.95	.07	6.86
		(1) $q_1 = .1, q_2 = .1$	5.41	5.02	.18	-1.98
		(2) $q_1 = .5, q_2 = .2 \begin{cases} (1) \\ (2) \end{cases}$	.07	.05	.39	-17.20
			14.48	14.46	.06	14.39

NOTE.— $Z_{TR}$  = expected maximum LOD score calculated using the true (generating) model;  $Z_{MG}$  = expected maximum LOD score calculated using the MG parameters;  $\theta$  = recombination estimate, using the MG parameters;  $Z_{MG}(\theta_0)$  = expected LOD score, using the MG parameters, for the true recombination fraction ( $= .05$ ).

<sup>a</sup> Results of LOD-score analysis with (1) marker A and with (2) marker B.

score appears when searching for linkage with these parameters for the recessive gene. These biases are non-negligible (LOD score divided by 2) and induce a failure to detect linkage.

There are models for which linkage is evidenced but for which the recombination fraction is overestimated. In addition, the true value of the recombination fraction may be strongly rejected. For example, for the two genes of model G1, linkage is concluded with  $Z_{MG1} = Z_{MG2} = 5.02$  for  $\theta = .18$ , but the true value of  $\theta = .05$  is

excluded with  $Z_{MG1}(\theta_0) = Z_{MG2}(\theta_0) = -2$ . Note that, in some cases,  $\theta$  may be underestimated as for the second gene of model C2.

There is thus an impact on detection of linkage due to the use of the MG parameters obtained by segregation analysis. But, in practice, it is impossible to know that these parameters are not correct. We therefore investigated two alternative approaches that would be helpful in such cases, the sib-pair method and the linkage-homogeneity test.

### Alternative Approaches to Detect Linkage

*Comparison of the power to detect linkage of the LOD-score method and the sib-pair method.*—In brief, the principle of the sib-pair method (Day and Simons 1976; Thomson and Bodmer 1977; Suarez 1978; Suarez et al. 1978) is to classify affected sib pairs according to the number of parental haplotypes they inherited in common, called "IBD" (identity by descent), which can take the value 2, 1, or 0. The test for linkage is performed by comparing the observed IBD distribution with the one expected under independence of segregation between disease and marker (.25, .5, .25). When the correspondence between the phenotypes and the genotypes is known, the power to detect linkage is greater by the LOD-score method than by the sib-pair method. This may not be true when the parameters are misspecified in the LOD-score analysis. Here we compared the power to detect linkage by the LOD-score method with the MG parameter estimates and by the sib-pair method. In  $N$  families generated under a two-locus model (as described in Methods), we calculated and maximized the expected LOD score  $N \times E[Z(\theta)]$ . We applied then the log-likelihood ratio test by transforming the maximum LOD score in  $\delta = 2 \times e \times \{\max N \times E[Z(\theta)]\}$ , which follows a  $\chi^2$  distribution with 1 df. Under the same two-locus model, we derived the IBD distribution expected in a sample of  $N$  affected sib pairs and calculated the  $\chi^2$  difference  $\delta'$  between this IBD distribution and the one equal to  $(.25 \times N, .50 \times N, .25 \times N)$ . We can then obtain the significance level of detection of linkage for both tests for a sample size  $N$ , which was set here to 100. For an easier comparison of both methods, we also calculated the sample sizes necessary to conclude for linkage with a significance level of 1/1,000 by setting the  $\delta$  and  $\delta'$  values to the critical values in the  $\chi^2$  distribution with 1 and 2 df, respectively, for the LOD-score method and the sib-pair method.

For models for which there is no bias or only a small one in the LOD-score estimation using the MG parameters, it is evident that the LOD-score method will be more efficient to detect linkage, as can be verified here. For example, when one considers the second gene of model E1, the  $Z_{MG2}$  value obtained by the LOD-score method using the MG parameter is not biased i.e., equal to  $Z_{TR2}$  (4.79;  $P = .3 \times 10^{-5}$ ) obtained using the true model. By the sib-pair method, we also concluded for linkage, but indeed with a lower significance level ( $\chi^2 = 14.26$ , 2 df;  $P = .8 \times 10^{-3}$ ).

Note that the above results are not included in table 3, where only models for which the bias in the LOD-score estimation is non-negligible for one of the genes are presented. This is the case for the first gene of models B2, E2, and F2. Table 3 shows the expected IBD distributions and the values of the  $\chi^2$  tests of the sib-pair

analysis and of the LOD-score analysis when the MG parameters are used. We see that, for these three models, the detection of linkage is more significant for the sib-pair method than for the LOD-score method when the MG parameters are used. For example, considering the first gene of model F2, we obtained, using the MG parameters, a  $Z_{MG1}$  value equal to 1.71 that corresponds to a  $P$  value of .005, while the  $P$  value decreased to .0002 when absence of linkage by the sib-pair method was rejected. In other words, with a significance level at 1/1,000 for conclusion of linkage, a sample size of 138 families would be necessary to conclude for linkage with the LOD-score method, while only 87 sib pairs would be necessary by the sib-pair method. The sib-pair method can thus lead to a non-negligible increase of power to detect linkage, when misspecification of parameters leads to a bias in estimation of the LOD score.

*Linkage-homogeneity test among subsamples.*—Maximization of the likelihood in a space of models including the underlying model leads to an estimation of the true parameters of the model. This result will hold in subsets of the data if correction for the selection criterion of the families is included in the analysis. Conversely, parameter estimates are not expected to be the same in each subset when the space of the model does not include the true model. As a consequence, testing homogeneity among subsamples is a way to indicate that the model specified is not the correct one. This approach has been described elsewhere (Dizier et al. 1992), and its principle initially illustrated by MacLean et al. (1975) in order to detect a spurious MG. We propose here to stratify the sample in subsamples  $i$  differing by the disease familial distribution and to apply the predivided sample test (Morton 1956) by using the MG parameters. The difference  $D = -2[\text{Ln}(L) - \sum_i \text{Ln}(L_i)]$  follows a  $\chi^2$  distribution with  $m(n - 1)$  df ( $m$  number of parameters and  $n$  number of subsamples), where  $L$  and  $L_i$  are the likelihood given the marker segregation conditional on the disease, respectively, in the whole sample and in the subsample  $i$ . Rejection of homogeneity indicates misspecification of the model. In addition, since heterogeneity is not expected in the absence of linkage, presence of heterogeneity also indicates the presence of linkage.

Using the family data generated under the different two-locus models as described in Methods, we applied the predivided sample test in subsamples divided according to status of the parents: none affected or at least one affected.

Linkage homogeneity may be rejected for the first gene of the three models B2, E2, and F2, with a respective  $\chi^2$  (1 df) equal to 4.63 ( $P = .03$ ), 6.07 ( $P = .014$ ), and 5.18 ( $P = .016$ ). This allows conclusion for both linkage and parameter misspecification. In fact, these three loci are those for which the impact of using MG parameters in LOD-score analysis was the most important.

**Table 3**  
**Comparison of Power to Detect Linkage by the Sib-Pair Method and by the LOD-Score Method, Using the MG Parameters**

GENERATING MODEL	PENETRANCE MATRIX	GENE FREQUENCIES	IBD DISTRIBUTION <sup>a</sup> EXPECTED IN 100 AFFECTED SIB PAIRS	TEST OF ABSENCE OF LINKAGE BY <sup>a</sup>	
				Sib-Pair Method <sup>b</sup>	LOD-Score Method <sup>b</sup>
Model B	$F = \begin{bmatrix} .5 & .5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	(2) $q_1 = .5, q_2 = .1$	(39.3, 49.4, 11.3)	$\chi^2$ 2 df = 14.4; $P = .0007$ ; $n_1 = 95$	$\chi^2$ 1 df = 9.9; $P = .005$ ; $n_2 = 109.4$
Model E	$F = \begin{bmatrix} .5 & .5 & .5 \\ .5 & .5 & 0 \\ .5 & .5 & 0 \end{bmatrix}$	(2) $q_1 = .2, q_2 = .01$	(44.3, 41.8, 13.9)	$\chi^2$ 2 df = 21.2; $P = .00003$ ; $n_1 = 65.2$	$\chi^2$ 1 df = 12.1; $P = .0005$ ; $n_2 = 89.5$
Model F	$F = \begin{bmatrix} .5 & .5 & 0 \\ .1 & 0 & 0 \\ .1 & 0 & 0 \end{bmatrix}$	(2) $q_1 = .5, q_2 = .05$	(40.3, 46.6, 13.3)	$\chi^2$ 2 df = 15.9; $P = .0002$ ; $n_1 = 86.9$	$\chi^2$ 1 df = 7.9; $P = .002$ ; $n_2 = 137.1$

<sup>a</sup> For the gene (2) of the generating model.

<sup>b</sup>  $n_1$  and  $n_2$  represent sample sizes necessary to detect linkage with a significance level of 1/1,000, with the sib-pair method and the LOD-score method, respectively.

## Discussion

The classical strategy, segregation analysis followed by LOD-score analysis, which is used to detect and localize a gene involved in a disease, was investigated here for the case of a trait depending on the interactive effect of two genes. Seeking evidence for linkage with genetic markers has often been justified by preliminary evidence of a MG effect by segregation analysis. In this case, screening the genome to detect linkage between the MG evidenced by segregation analysis and a marker would seem justified. However, our study shows that strong evidence of a MG effect by segregation analysis obtained for some two-locus models does not necessarily imply that linkage will be detected for at least one of the genes, even with a close informative marker and with the true parameters. Indeed, there appears to be no clear correspondence between the power to detect a MG by segregation analysis and the power to detect linkage by the LOD-score analysis.

For two-locus models for which evidence of linkage was obtained for at least one gene when the true parameters of this gene were used, it was then important to investigate the impact of using the MG parameters obtained by segregation analysis instead of the true parameters. Previous studies were already performed to evaluate the impact of ignoring the effect of a second gene (Goldin 1992; Vieland et al. 1992). They found no significant difference when a single or a two-locus model was used in the LOD-score analysis; however, the authors used the correct single-locus model or at least the correct transmission mode in the analysis. In contrast, we found here that, for some two-locus models, use of the MG parameters instead of the true parameters may affect the detection of linkage with bias in the estimation of the LOD score. This is particularly true when the two genes have different characteristics (allelic frequency and mode of transmission). Note that this type of model may be, in reality, more frequent than in our panel of examples. On the other hand, for most models (symmetrical or not), there is overestimation of the recombination fraction and even in some cases an exclusion for the true recombination fraction while linkage has been concluded for another region. This would be a crucial problem for the further investigation by chromosome walking to identify the susceptibility gene.

The difficulties of parameterizing the gene effect in complex disorders led to an increased interest for non-parametric approaches to seek for linkage with genetic markers. For diseases depending on the effect of two genes, Goldin and Weeks (1993) compared the power to detect linkage of the LOD-score method and of non-parametric methods including the sib-pair method. They found a greater power of the LOD-score method when

using the true parameters of the gene in the analysis. However, as noted by the authors, when the parameters are not known, different models have to be used, in particular different modes of transmission: this leads to increased degrees of freedom of the test and decreased power to detect linkage. We show here that, when compared to the LOD-score method using the MG parameters obtained by segregation analysis, the sib-pair method may give a non-negligible increase of power to detect linkage, particularly when the misspecification of the parameters leads to an important bias in the LOD-score estimation and a failure to detect linkage. For these models we may also apply the homogeneity sample test. However, the power to detect linkage is lower than with the sib-pair method. The main interest of such an approach is that it simultaneously provides evidence for linkage and misspecification of the parameters.

In conclusion, even when segregation analysis has been conclusive for a MG effect, use of the estimated MG parameters may affect conclusions of LOD-score analysis. This is illustrated here for the case of two-locus models. We also showed the interest of using other approaches as the sib-pair method and the linkage-homogeneity test, which appear very complementary to the LOD-score method for studying complex traits.

## Acknowledgments

We would like to thank R. Spielman and an anonymous reviewer for their very helpful comments.

## References

- Clerget-Darpoux F, Bonaiti-Pellié C, Hochez J (1986) Effects of misspecifying genetic parameters in LOD score analysis. *Biometrics* 42:393–399
- Day NE, Simons JM (1976) Disease susceptibility genes: their identification by multiple case family studies. *Tissue Antigens* 8:109–119
- Dizier MH, Bonaiti-Pellié C, Clerget-Darpoux F (1993) Conclusions of segregation analysis for family data generated under two-locus models. *Am J Hum Genet* 53:1338–1346
- Dizier MH, Clerget-Darpoux F, Hochez J (1992) Testing genetic models through subsample homogeneity. *Life Sci Adv* 11:215–221
- Elston RC, Stewart J (1971) A general model for the analysis of pedigree data. *Hum Hered* 21:253–242
- Goldin LR (1992) Detection of linkage under heterogeneity: comparison of the two-locus vs admixture models. *Genet Epidemiol* 9:61–66
- Goldin LR, Weeks DE (1993) Two-locus models of disease: comparison of likelihood and nonparametric linkage methods. *Am J Hum Genet* 53:908–915
- Lalouel JM (1979) A computer program for optimization of general nonlinear function. Tech rep 14, Department of



- Medical Biophysics and Computing, University of Utah, Salt Lake City
- Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. *Am J Hum Genet* 35:816-826
- MacLean CJ, Morton NE, Lew R (1975) Analysis of family resemblance. IV. Operational characteristics of segregation analysis. *Am J Hum Genet* 27:365-384
- Morton NE (1955) Sequential tests for detection of linkage. *Am J Hum Genet* 7:277-318
- (1956) The detection and estimation of linkage between the genes for elliptocytosis and the RH blood type. *Am J Hum Genet* 8:80-96
- Morton NE, Maclean CJ (1974) Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am J Hum Genet* 26:489-503
- Suarez BK (1978) The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens* 12:87-93
- 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
- Thomson G, Bodmer W (1977) The genetic analysis of HLA and disease association. In: Dausset J, Svejgaard A (eds) *HLA and disease*. Munksgaard, Copenhagen, pp 84-93
- Vieland VJ, Hodge SE, Greenberg DA (1992) Adequacy of single-locus approximations for linkage analysis of oligogenic traits. *Genet Epidemiol* 9:45-59