Ann. Hum. Genet. (2001), **65**, 473–482
Printed in Great Britain

A novel method of two-locus linkage analysis applied to a genome scan for late onset Alzheimer's disease

D. CURTIS[1], B. V. NORTH[1] AND P. C. SHAM[2]
[1] *Joint Academic Department of Psychological Medicine, St Bartholomew's and Royal London School of Medicine and Dentistry, 3rd Floor Alexandra Wing, Turner Street, London E1 1BB, UK*
[2] *Department of Psychological Medicine, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK*

(Received 14.3.01. Accepted 6.6.01)

SUMMARY

A number of methods have previously been described which carry out linkage analysis considering information for two or more loci simultaneously. Apart from some *ad hoc* methods such as analysing subsamples, these methods use information regarding linkage at all loci under consideration. However, if the actual genotype-specific effects are known for some loci then it would be preferable to consider the genotypes of these loci directly, rather than the amount of allele-sharing they demonstrate. Here we present an extension to our likelihood-based method of model-free linkage analysis as implemented in the *MFLINK* program. This allows the incorporation of liability classes. The genotypes of a locus known to affect risk can be used to assign subjects to liability classes prior to carrying out linkage tests at other loci. An example application is presented for genome scan data on Alzheimer's disease with analysis conditional on Apoliprotein E (*APOE*) genotypes. The results provide support for the existence of additional susceptibility loci linked to D10S1211 and to D12S358.

INTRODUCTION

When variation at more than one locus affects susceptibility to disease it is reasonable to assume that considering the effects of such loci simultaneously may enhance the performance of linkage analysis. Such approaches may be increasingly important as attention focuses on the study of traits with complex inheritance, where it is expected that multiple loci may be involved. A number of methods have been described previously. Although in principle they are readily extensible to consideration of three or more loci, in practice such methods have generally been applied to two susceptibility loci. Likewise, although some approaches can deal with the analysis of two linked susceptibility loci, the theory and implementation is usually much simpler if it is assumed that the two loci are unlinked.

One approach is a conceptually straightforward extension of the classical lod score method to incorporate two affection loci exerting a joint effect on risk. This method has been implemented in the TLINK programs (Schork *et al.* 1993), and has been applied to the analysis of multiple sclerosis (Tienari *et al.* 1994). If both loci are biallelic then the relationship between genotype at the affection loci and risk of affection is modelled by a 3×3 table of penetrance probabilities, each of which represents the probability of affection conditional on the joint genotype at the two loci. Standard likelihood calculations are carried out considering the affection status of subjects in pedigrees and their genotypes at markers linked to the affection loci. Linkage parameters are estimated according to the likelihoods derived from setting different recom-

Correspondence: D. Curtis.
E-mail: dcurtis@hgmp.mrc.ac.uk

bination fractions between each affection locus and the marker or markers hypothetically linked to it. In order to carry out this kind of two-locus linkage analysis it is necessary to specify the full table of penetrance values that define the transmission model, along with allele frequencies at both affection loci. It is perhaps also worth mentioning that the likelihood calculations can be quite demanding of computer time and memory, since they require simultaneous consideration of at least four loci (two affection loci and two marker loci), although a faster Markov chain Monte Carlo (MCMC) method of estimating the two-locus lod score has recently been described (Lin, 2000). One special instance of two-locus likelihood-based linkage analysis which is computationally far less demanding is admixture analysis. Here the assumption is that in any single pedigree only one or the other (or neither) locus will be exerting an effect on risk and that interactions between the loci do not need to be considered. In this situation one can carry out likelihood calculations assuming linkage or non-linkage at each of the loci separately, and then combine the likelihoods in a simple extension to the standard test for linkage assuming admixture (Risch, 1989), in order to test for one or other or both loci demonstrating linkage. The method is simplified if it is assumed that both loci have the same effect on risk, i.e. have the same transmission model, though this is not an essential feature of the approach. Such a method is implemented in Ott's HOMOG3 program (Ott, 1991) and has, for example, been applied to the analysis of bipolar affective disorder (Smyth *et al*. 1997).

A disadvantage of the two-locus lod score methods is that transmission models need to be specified by the user. In certain circumstances this may not be problematic. For example if one has a number of moderately-sized pedigrees segregating a Mendelian dominant disease then they might appropriately be analysed using a two-locus admixture analysis. This would allow for the possibility that individual pedigrees may not be large enough to be definitely classified as linked or unlinked to either locus. However it will often be the case that the true transmission model is unknown. This will especially be so where two loci interact, because even if one can estimate overall segregation parameters for the disease these will not provide information regarding the relative contributions of the loci. The affected sib pair (ASP) method of linkage analysis represents an example of such a model-free approach and has been extended to two-locus analysis (Dizier & Clerget-Darpoux, 1986).

A useful distinction may be drawn between the situation where one is attempting to implicate two loci simultaneously, as opposed to the situation where one locus is already well established and one is investigating whether a second locus is also involved. Attempting to implicate two loci at once produces formidable problems of multiple-testing, which may mean that such approaches have little if any advantage over single locus methods. A commonly recommended approach is to apply two-locus methods only to markers showing some suggestion of linkage in single locus analyses. However, if a pair of markers by chance produces support for linkage in a two-locus analysis then it is likely that they will also be somewhat positive in the single locus analyses, so the approach of *post hoc* selection does not really avoid the problem that from 300 markers one can select any of about 45000 pairs of markers, apparently necessitating an appropriate correction for the implicit multiple testing involved. A different situation arises when one locus is already known to affect susceptibility and one is seeking to discover further loci. In such a case it would seem advantageous to incorporate information from the first locus when searching for others which may also exert an effect. Methods that condition on information from a previously known locus have been described for sib-pairs (Cordell *et al*. 1995) and more recently for general pedigrees (Cordell *et al*. 2000).

It is also helpful to distinguish between the situation where linkage is well-established at the first locus, from that where the allele-specific effects of a particular polymorphic system influencing susceptibility have been characterised.

In the former situation, the only information which can be used in the search for additional susceptibility loci is the extent of linkage manifest at the first locus, commonly characterised as the extent of excess allele-sharing between affected subjects. However, if one knows which particular polymorphism influences risk of affection then one can characterise the nature and magnitude of this effect through epidemiological studies. It would then make sense to incorporate information regarding which alleles of this system each subject possessed when attempting to elucidate contributions from other loci. We can regard this as a linkage analysis at the second (unknown) locus that is conditioned by the genotypes at the first (known) locus, rather than being conditioned only on linkage information. From a computational point of view, it is easier to carry out such a linkage analysis conditional on genotypes, because calculations to obtain linkage parameters at the first locus are unnecessary. Rather, the genotypes at the first locus will in some way be used to inform the analysis at the second locus. One approach to this is to carry out stratified analyses, in which different groups of subjects are analysed separately according to their genotypes at the first locus. Such approaches have been applied to diabetes (Mein *et al.* 1998) and Alzheimer's disease (AD) (Kehoe *et al.* 1999). However, as Cordell and colleagues point out (Cordell *et al.* 2000), there can be problems in deciding how to classify subjects, and the multiple subsamples may necessitate a Bonferroni correction which could decrease overall power. Additionally, although subdivision may be relatively straightforward for small units of data such as ASPs, difficulties arise if one attempts to use general pedigrees, when members of the same pedigree may have several different genotypes at the first locus.

A more satisfactory approach would be to carry out a single analysis which combined information regarding genotypes at the first locus with linkage at the second. If the nature of the effects at the two loci were known, it would be trivial to implement such a method using classical lod score analysis. One would simply use the genotypes at the first locus to assign subjects to one of a number of liability classes (three, assuming a biallelic locus). Then different sets of penetrance values for the second locus would be specified for each of these liability classes and a standard lod score analysis would be performed. However, as already stated, the true interactive transmission model is unlikely to be known, rendering this approach problematic. Here we present an extension to our previously-described method of model-free linkage analysis in order to test a particular class of combined transmission models, that in which the effect of the first locus on risk has a constant multiplicative effect for all genotypes at the second locus.

METHODS

We have previously presented a method of linkage analysis based on likelihoods which is model-free in the sense that a transmission model does not need to specified *a priori* (Curtis & Sham, 1995). A single position on the genetic map is tested, meaning that the recombination fractions between the putative affection locus and the markers are not varied. A population prevalence K for the trait is set, and then likelihoods are calculated for a range of different transmission models, each of which would yield this overall prevalence. The models tested range from Mendelian dominant through null effect (all penetrances equal) to Mendelian recessive. For each model, likelihoods are calculated assuming that no families within the sample are linked, that all are linked or that a proportion are linked. Three different lod scores are derived from these likelihoods. The MLOD is the maximum difference in log likelihoods between linkage and non-linkage for any model, the MALOD is the maximum difference between admixture and non-linkage for any model, and the MFLOD is the difference between the maximum likelihood for any model under admixture and the maximum likelihood, possibly for a different model, under non-linkage. We have shown theoretically and empirically that the associated likelihood

ratio test statistics (i.e. 2lnLR = 2ln(10)lod) are approximately distributed under the null hypothesis as $X^2_{(1)}$, a 50:50 mixture of $X^2_{(2)}$ and $X^2_{(0)}$ and a 50:50 mixture of $X^2_{(1)}$ and $X^2_{(0)}$ respectively (Curtis *et al.* 1999; Sham *et al.* 2000). This method has been implemented in the MFLINK program.

The key to applying MFLINK to a two-locus analysis where linkage analysis at the second locus is conditional on the genotypes at the first is to extend MFLINK to deal with liability classes. In the original version of MFLINK, each transmission model is specified by three penetrance values, denoted f_0, f_1 and f_2, which denote the risk of affection conditional on genotype AA, Aa or aa at the affection locus. The prevalence K of the trait is fixed, which implies that the frequency of the susceptibility allele can be calculated once these penetrance values are specified. The range of transmission models tested is defined by linearly varying these values from $f_0 = 0, f_1 = f_2 = 1$ (Mendelian dominant) to $f_0 = f_1 = f_2 = K$ (null effect) and then to $f_0 = f_1 = 0$, $f_2 = 1$ (Mendelian recessive). To use liability classes, we write f_{ij} to indicate the ith penetrance value ($i = 0$, 1 or 2) for the jth liability class. The overall prevalence of affection among subjects in the jth liability class is written K_j. There will be one liability class for which this prevalence is maximal and we write this prevalence as K_{MAX} with the penetrance values for this class written as $f_{i\text{MAX}}$.

Extending the MFLINK method to liability classes depends on having a method for defining an appropriate range of transmission models whose likelihoods are to be calculated. Various schemes could be chosen. The one we propose here is that for those subjects falling into the liability class with penetrance K_{MAX} the transmission model is varied between Mendelian dominant, null effect and Mendelian recessive while for those in other liability classes the penetrance is always defined as $f_{ij} = f_{i\text{MAX}} * K_j/K_{\text{MAX}}$. What this means is that for each genotype AA, Aa or aa the ratio of the penetrance values between subjects in different liability classes is kept constant. For subjects not in the class with maximal prevalence, the transmission models tested do not range up to the Mendelian ones but to models with penetrance values of K_j/K_{MAX}.

To generate the range of models to be tested, $f_{1\text{MAX}}$ is varied from 0 to 1. Other penetrances are then defined as functions of $f_{1\text{MAX}}$ as follows:

For $f_{1\text{MAX}} < K_{\text{MAX}}: f_{0\text{MAX}} = f_{1\text{MAX}}$,

$$f_{2\text{MAX}} = f_{1\text{MAX}}(K_{\text{MAX}} - 1)/K_{\text{MAX}} + 1$$

Otherwise $f_{2\text{MAX}} = f_{1\text{MAX}}$,

$$f_{0\text{MAX}} = (1 - f_{1\text{MAX}})K_{\text{MAX}}/(1 - K_{\text{MAX}})$$

For all $f_{ij}, j\,! = MAX: f_{ij} = f_{i\text{MAX}} * K_j/K_{\text{MAX}}$

For each model, the correct frequency for the allele conferring increased susceptibility, q, is calculated by solving the quadratic equation:

$$K_{\text{MAX}} = f_{0\text{MAX}}(1-q)^2 + f_{1\text{MAX}}2q(1-q) + f_{2\text{MAX}}q^2$$

(The same value for the allele frequency will yield correct prevalences for the other liability classes, $K_{j\,!=\text{MAX}}$.)

A relatively small number of transmission models is tested, with the expectation that one will be close enough to the true model to detect any evidence for linkage which may be present. Typically, one might use 5 values of $f_{1\text{MAX}}$ below K_{MAX}, K_{MAX} itself (the null model), and 5 values above K_{MAX}. If we write $\theta = t$ to imply the disease locus is at the test position (which may be on a multipoint map), $\theta = 0.5$ to indicate it is unlinked to any markers and α to be the proportion of families in which the disease locus exerts an effect, then for each model we calculate three likelihoods which are functions of $f_{1\text{MAX}}$, θ and α:

$$L_{\text{UNLINKED}} = L(\theta = 0.5 \text{ or } \alpha = 0, f_{1\text{MAX}})$$

$$L_{\text{LINKED}} = L(\theta = t, \alpha = 1, f_{1\text{MAX}})$$

$$L_{\text{HET}} = L(\theta = t, \alpha, f_{1\text{MAX}})$$

where L_{HET} is maximised over α

The three lod scores are then defined as:

$$\text{MLOD} = \log_{10}(\max[L_{\text{LINKED}}/L_{\text{UNLINKED}}])$$

$$\text{MALOD} = \log_{10}(\max[L_{\text{HET}}/L_{\text{UNLINKED}}])$$

$$\text{MFLOD} = \log_{10}(\max[L_{\text{HET}}]/\max[L_{\text{UNLINKED}}])$$

Their distributions under the null hypothesis are related to chi-squared distributions as noted above.

Once we have established a scheme for extending MFLINK to deal with liability classes, we can use it to implement a two-locus method of linkage analysis where the genotype-specific effects of one locus are already established. We use each genotype of the first locus to define a liability class for linkage analysis at the second locus, with the prevalence for the trait K_j being set to the value appropriate for that genotype, which may well be known with some degree of accuracy from epidemiological studies. If the first locus is biallelic then there will be three liability classes corresponding to the three possible genotypes. The method assumes that the genotype at the first locus has a constant multiplicative effect on risk for each genotype of the second locus, i.e. that the ratios $f_{i0}:f_{i1}:f_{i2}$ are the same for each genotype i ($i = 0, 1, 2$) of the second locus.

Example application

To clarify the procedure and to provide an example of its performance in practice, we present here an application to genome scan data for Alzheimer's disease (AD). The dataset was obtained from the NIMH Alzheimer Disease Genetics Initiative Data Archive (*http://zork.wustl.edu/nimh/ad.html*) and 292 ASPs from these pedigrees were previously subject to a first-stage genome scan (Kehoe *et al.* 1999). Here, we include in the analysis all pedigree members rather than only affected sib pairs. The distributed dataset contains 241 pedigrees typed for 237 microsatellite markers and for the gene for Apoliprotein E (*APOE*). Six pedigrees consist simply of an affected subject with two parents. These would be suitable for TDT analysis but provide no information for linkage analysis. Nevertheless, for reasons of convenience they were included in the analysis and would have no effect on its results. The default definition of affection as contained in the distribution files was used, consisting of 'definite' or 'probable' AD as defined for the affected sib pair scan. Most pedigree members were old enough to have had the possibility to develop AD, but a few were below the age of 60. For purposes of linkage analysis, it would make sense to encode such young unaffected subjects as being of unknown affection status rather than as being unaffected, or alternatively to implement age-related liability classes. Once again, this was not done for reasons of convenience and the supplied pedigree files were used to assign affection status without modification. The prime purpose of the present exercise is to provide an example comparison of the results of single-locus and two-locus analysis, rather than to perform an exhaustive analysis of the data, which potentially could involve many additional subtleties and would probably necessitate additional genotyping at areas of interest. A handful of genotypes were found to be incompatible or incorrectly coded, and these were omitted.

Two methods of analysis were applied, a standard single locus MFLINK analysis and the new method in which liability classes are defined conditional on the genotype at a known susceptibility locus, in this case *APOE*. Two-point analyses were carried out with each marker with a test position set at a recombination fraction of 0.05 with the marker. Three-point analyses were carried out using pairs of consecutive markers with the test position set midway between the markers and at a recombination fraction of 0.05 with the first and last marker on each chromosome. In order to approximately model the known effects on risk of *APOE*, we considered only the effects of the $\epsilon 4$ allele (even though it is known that $\epsilon 2$ confers somewhat lower risk than $\epsilon 3$ (Corder *et al.* 1994)) and hence only the genotypes $\epsilon 4/\epsilon 4$, $\epsilon 4$/* and */* were used to define three liability classes. We assigned subjects to one of these classes based on this *APOE* genotype, and included a fourth liability class for subjects whose *APOE* type was unknown. We assumed that the $\epsilon 4$ allele had an approximate prevalence of 0.2, and that each $\epsilon 4$ allele increased risk by a factor of approximately 3 in a multiplicative fashion. These considerations led us to assign prevalences for each liability

Table 1. *Results from two-point analyses are shown in line with the marker concerned, while results of three-point analyses are shown between markers*

		Standard MFLINK analysis			Two-locus method		
Marker	Position	MLOD	MALOD	MFLOD	MLOD	MALOD	MFLOD
D1S1631	108.3	0.4797	0.4797	0.1678	0.2651	0.2651	0.1049
	111.0	0.8498	0.8720	0.5196	0.4633	0.4633	0.2463
D1S1675	113.8	0.4232	0.4232	0.0613	0.1012	0.1012	0.0117
D2S427	161.8	0.5152	1.0831	0.5337	0.0713	0.0713	0.0000
	171.3	0.2729	0.4211	0.4211	0.0134	0.0134	0.0000
D2S125	180.4	0.7022	0.8188	0.7535	0.2140	0.2140	0.0127
	185.4	0.5420	0.7138	0.7138	0.0905	0.0905	0.0000
D3S1303	137.0	0.6756	1.4840	0.9106	0.4596	0.4596	0.3339
D3S1314	209.7	0.3706	0.6333	0.6333	0.1524	0.1542	0.0616
D5S2845	18.9	0.2517	1.4594	0.5323	0.9328	0.9328	0.9328
	32.5	0.7210	1.6079	1.0640	1.1063	1.1063	1.1063
D5S1470	46.1	0.5664	0.7703	0.6255	0.6703	0.6703	0.5674
	57.9	0.8562	1.1548	0.8141	1.1048	1.1048	1.0959
D5S1457	69.8	0.4700	0.6514	0.4685	0.5229	0.5229	0.4397
D6S1018	44.5	0.0000	0.1769	0.0193	0.0030	0.0047	0.0000
	67.5	0.2244	0.6188	0.6188	0.1505	0.1505	0.0754
D6S1036	70.5	0.2920	0.8077	0.8077	0.2917	0.2919	0.1825
D6S1021	119.7	0.6509	0.8183	0.7705	0.6494	0.6817	0.5617
	121.7	0.8003	1.5667	1.5667	0.7622	0.7786	0.5592
D6S474	123.8	0.6774	1.4608	1.4608	0.5491	0.5491	0.3496
	141.5	0.0755	0.8602	0.8602	0.2044	0.2151	0.1145
D6S495	159.3	0.0000	0.0459	0.0459	0.0000	0.0493	0.0243
D8S1102	54.7	0.0446	0.0508	0.0000	0.5556	0.5678	0.5562
	62.7	0.1701	0.3097	0.0000	0.9183	1.1377	0.9231
GA12B06	68.6	0.5536	1.0765	0.2201	1.2175	1.3017	1.0424
		0.6201	0.7988	0.0946	0.9534	1.0486	0.8995
D8S1119	105.1	0.3001	0.3555	0.0000	0.2804	0.2855	0.2357
D9S285	13.8	0.1942	0.1942	0.0000	0.0932	0.0965	0.0706
	16.8	0.9426	1.0399	0.4513	0.5420	0.5420	0.4404
D9S171	19.9	1.2943	1.3606	0.9354	0.7259	0.7259	0.6682
D9S176	103.3	1.5697	1.5697	0.6154	0.9636	0.9636	0.6871
D10S1426	23.4	0.3546	0.3546	0.3100	0.1780	0.1780	0.1745
	45.9	1.7497	1.7497	1.3949	1.2893	1.2893	1.2333
D10S1211	68.5	2.0410	2.0410	1.3924	1.9415	1.9415	1.7756
	72.6	1.4796	1.7312	1.4893	1.9151	1.9151	1.5873
D10S676	78.7	0.3159	0.6611	0.6443	0.5723	0.6330	0.4662
	93.3	0.2343	0.7568	0.7568	0.6785	0.6785	0.5383
D10S670	107.9	0.0482	0.2536	0.2536	0.1259	0.1278	0.0971
D11S2002	102.0	0.7795	0.9504	0.7381	0.6505	0.6505	0.3683
	103.3	0.8716	0.9392	0.7591	0.6158	0.6158	0.3715
D11S1354	104.6	0.0000	0.0002	0.0000	0.0000	0.0001	0.0000
D12S397	17.7	0.5458	0.7933	0.7933	0.3459	0.3459	0.1942
		0.4788	0.7475	0.7475	0.3989	0.4088	0.3800
D12S98	24.5	0.3955	0.4073	0.3302	0.4267	0.4286	0.4286
		1.3292	1.5423	1.3389	2.2181	2.2181	2.2181
D12S358	26.2	1.8427	2.0931	2.0931	2.1178	2.1178	2.0911
D17S2193	35.0	1.0990	1.2153	0.4198	0.4873	0.4873	0.0898
	44.6	0.9303	0.9335	0.0322	0.6466	0.6466	0.3113
D17S1290	54.2	0.6097	0.6508	0.0000	0.8643	0.8643	0.6157
D19S225	40.6	0.4108	0.5140	0.0000	0.1546	0.1546	0.0539
		1.2155	1.2433	0.3596	0.4292	0.4292	0.1860
D19S412	55.4	0.8630	0.9868	0.2363	0.3187	0.3187	0.0571
	57.6	1.9314	1.9433	1.3436	0.6211	0.6211	0.2106
D19S571	59.9	1.7567	1.7803	1.6134	0.5650	0.5650	0.3013
		0.8818	0.9882	0.9599	0.4105	0.4105	0.2378
D19S210	62.1	0.1276	0.1323	0.0562	0.1223	0.1260	0.1165

Table 1. (Contd....)

Marker	Position	Standard MFLINK analysis MLOD	MALOD	MFLOD	Two-locus method MLOD	MALOD	MFLOD
D20S120	60.9	0.6636	0.8902	0.6297	0.3382	0.3382	0.0657
D22S419	18.8	0.0453	0.7066	0.7066	0.1710	0.3673	0.2483
DXS1002	86.0	0.3625	0.6506	0.6506	0.4483	0.4483	0.2814

The table shows all regions for which at least one result is nominally significant at $p < 0.05$. Results from two-point analyses are shown in line with the marker concerned, while results of three-point analyses are shown between markers.
Map position is given in centimorgans from *pter* according to the genetic location database (LDB, *cedar.genetics.soton.ac.uk/public_html/ldb.html*) (Collins *et al.* 1996), except for the chromosome 12 markers where positions from the Marshfield map (*research.marshfieldclinic.org/genetics*) obtained from UDB (Chalifa-Caspi *et al.* 2000) are given. For markers absent from LDB, nearby markers were used to deduce an approximate position.

Table 2. *Table showing transmission model parameters producing maximum admixture lod scores for A) D12S358 and B) D19S571*

A) Penetrance values for D12S358

		Genotype at test locus, Frequency of allele a = 0.77, alpha = 1.0: AA	Aa	aa
Apolipoprotein E genotype	$\epsilon4/\epsilon4$	0.001	0.001	0.999
Frequency of allele $\epsilon4 = 0.2$	$\epsilon4/*$	0.00036	0.00036	0.36
	/	0.00011	0.00011	0.11

B) Penetrance values for D19S571

		Genotype at test locus, Frequency of allele a = 0.36, alpha = 1.0: AA	Aa	aa
Apolipoprotein E genotype	$\epsilon4/\epsilon4$	0.24	0.84	0.84
Frequency of allele $\epsilon4 = 0.2$	$\epsilon4/*$	0.087	0.30	0.30
	/	0.026	0.091	0.091

class of 0.59, 0.20 and 0.065 respectively. Assuming an $\epsilon4$ allele frequency of 0.2 this would produce an overall population frequency for AD of 0.13 among elderly subjects, which is somewhat on the high side but good enough for the current purpose. Minor changes in such parameters have negligible effects on the results of linkage analysis (Clerget-Darpoux *et al.* 1986). The prevalence of 0.13 was used for the liability class in which *APOE* status was unknown, and also for the single locus analyses.

RESULTS

Table 1 shows the results for all markers in which any of the three lod scores produced by either method was significant at $p < 0.05$. Purely to provide examples of some of the transmission models which would be tested by MFLINK, Table 2 shows the parameters of the two-locus models which produced maximum admixture lod scores (MALOD) for D12S358 and D19S571.

Although there are no situations in which the two methods give radically different results, for several markers (D2S247, D3S1303, D5S2845, D6S474, D9S176, D17S2193) there is a modestly positive result from the standard method, but somewhat or greatly reduced support for linkage from the two-locus method. By contrast, two regions provide fairly strong support for linkage using both methods of analysis. D10S1211 produces an MLOD of 2.04 using the standard method and an MFLOD of 1.78 using the two-locus method, both results have nominal significance of $p = 0.002$. D12S358 produces an

MFLOD of 2.09 using both methods, having a nominal significance of $p = 0.001$. The three-point analysis using D12S98 and D12S358 produces a very slightly more significant result with the two-locus method, consisting of an MFLOD of 2.2 ($p = 0.0007$). For one other marker, GA12B06, the support for linkage is slightly stronger with the second method: the standard method produces an MALOD of 1.08 ($p = 0.04$) while the two-locus method produces an MFLOD of 1.04 ($p = 0.01$).

DISCUSSION

The method described represents a simple extension to the single locus method of model-free linkage analysis previously implemented in MFLINK. Unlike other two-locus methods, it can incorporate genotype, rather than linkage, information from an established locus when searching for additional susceptibility loci. It should therefore be increasingly useful as risk loci are identified for diseases with complex modes of transmission.

Quantifying the advantages of the new method will depend on extensive studies of simulated and real data. It seems likely that, depending on the exact nature of the interactions between loci, in some circumstances it may offer substantial benefits while in others very little. The method as described assumes that the first locus has a constant multiplicative effect on risk for each genotype at the second locus. Of course, this assumption may not be correct and it is not clear how sensitive the method is to assumptions regarding the nature of the interaction. With the exception of examples of admixture, to date there is insufficient information regarding known genetic interactions to judge what kind effects are likely to be commonly found in reality. Other types of interaction could be modelled according to similar principles and devising and testing such methods will be the subject of further work.

In the example application to the AD data it is perhaps fair to say that there is no real dramatic difference between the results obtained by the two methods. Nevertheless, the results of the two-locus method would serve to focus attention somewhat differently when considering what follow-up studies should be pursued. For a number of markers the evidence for linkage seems to diminish or disappear when the known effect of *APOE* is taken into account. If one believes that the approach is valid then this might lead one to suspect that the positive results from the standard analysis had arisen by chance and were less likely to indicate a true linkage. Alternatively, for the positive markers on chromosome 19 it seems probable that the positive results occur due to the effect of *APOE*, which lies only a few centimorgans away from them. Once the effect of *APOE* is taken account of in the two-locus analysis there is no residual evidence in support of linkage. (It is perhaps interesting to note that the standard analysis did produce an MLOD of 1.8, $p = 0.004$, indicating that this linkage study would have directed attention to the region containing *APOE* if its role had not already been identified.) For GA12B06 the evidence in favour of linkage is slightly increased when the new two-locus method is applied. Regions on chromosomes 10 and 12 are positive using both methods and produce as much evidence for linkage as the region around *APOE* itself, perhaps suggesting that they would be worthy of further investigation.

In the original ASP analysis (Kehoe *et al.* 1999) the region of chromosome 10 which we found to be positive produced a multipoint MLS score of 2.27 (simulation p value approx. 0.001). When the ASPs were divided according to $\epsilon 4$ status, those which were both $\epsilon 4$ positive produced an MLS of 2.25 while those which were both negative produced an MLS of 0.92, with essentially equal IBD proportions. Subsequently, additional ASPs were included and additional markers typed (not available for the current analysis) and the MLS rose to 3.83 in the second stage analysis (Myers *et al.* 2000). This same region has also demonstrated linkage to plasma amyloid $\beta 42$ peptide (A$\beta 42$) in a different sample of families selected through an AD proband with extremely high plasma levels of A$\beta 42$ (Ertekin-Taner *et al.* 2000). Another linkage study of AD

also detected linkage to 10q markers with a lod of 3.3, but at a location some 40 cM distal to the region implicated by the former studies (Bertram *et al*. 2000).

In the original ASP analysis (Kehoe *et al*. 1999) the region of chromosome 12 which produced positive results in our analyses produced a multipoint MLS score of 0.89 (simulation p value > 0.025), although when the sample was subdivided those sib pairs which were both $\epsilon 4$-negative produced an MLS of 1.9. Thus our own analysis produces somewhat stronger evidence for linkage than did the ASP approach. However linkage to this region has not as yet been reported in other studies.

The new model-free linkage analysis method conditional on genotypes at established risk loci is fairly straightforward to apply and has some intuitive appeal. The magnitude and nature of any advantage it may have over other methods will require further investigation. Nevertheless it seems that this, and similar approaches, offer some promise of increasing the power to identify additional susceptibility loci and could usefully be applied to real data.

Software availability

MFLINK and the accompanying MFMAP utility have been modified to allow analysis of affection loci having more than one liability class. An additional program, MAKELIAB, automatically assigns liability classes to one affection locus based on the genotype at another locus. All programs are provided as C source code and MSDOS executables in the MFLINK package obtainable from *www.mds.qmw.ac.uk/statgen/dcurtis/software.html*.

BVN is supported by Wellcome Trust Project Grant No. 055379. Data and biomaterials were collected in three projects that participated in the National Institute of Mental Health (NIMH) Alzheimer Disease Genetics Initiative. From 1991–98, the Principal Investigators and Co-Investigators were: Massachusetts General Hospital, Boston, MA, U01 MH46281, Marilyn S. Albert, Ph.D., and Deborah Blacker, M.D., Sc.D.; Johns Hopkins University, Baltimore, MD, U01 MH46290, Susan S. Bassett, Ph.D., Gary A. Chase, Ph.D., and Marshal F. Folstein, M.D.; University of Alabama, Birmingham, AL, U01 MH46373, Rodney C.P. Go, Ph.D., and Lindy E. Harrell, M.D. The authors are grateful to the two anonymous referees who provided comments which significantly improved the quality of this paper.

REFERENCES

Bertram, L., Blacker, D., Mullin, K., Keeney, D., Jones, J., Basu, S., Yhu, S., McInnis, M. G., Go, R. C., Vekrellis, K., Selkoe, D. J., Saunders, A. J. & Tanzi, R. E. (2000). Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. *Science* **290**, 2302–3.

Chalifa-Caspi, V., Prilusky, J. & Lancet, D. (2000). *The Unified Database* (*UDB*) *http://bioinformatics.weizmann.ac.il/udb*. Rehovot, Weizmann Institute of Science.

Clerget-Darpoux, F., Bonaiti-Pellie, C. & Hochez, J. (1986). Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* **42**, 393–9.

Collins, A., Frezal, J., Teague, J. & Morton, N. E. (1996). A metric map of humans: 23,500 loci in 850 bands. *Proc. Natl. Acad. Sci. USA* **93**, 14771–5.

Cordell H. J., Todd, J. A., Bennett, S. T., Kawaguchi, Y. & Farrall, M. (1995). Two-locus maximum lod score analysis of a multifactorial trait: joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes. *Am. J. Hum. Genet.* **57**, 920–34.

Cordell, H. J., Wedig, G. C., Jacobs, K. B. & Elston, R. C. (2000). Multilocus linkage tests based on affected relative pairs. *Am. J. Hum. Genet.* **66**, 1273–86.

Corder, E. H., Saunders, A. M., Risch, N. J., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Rimmler, J. B., Locke, P. A., Conneally, P. M., Schmader, K. E. *et al*. (1994). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat. Genet.* **7**, 180–4.

Curtis, D. & Sham, P. C. (1995). Model-free linkage analysis using likelihoods. *Am. J. Hum. Genet.* **57**, 703–16.

Curtis, D., Zhao, J. & Sham, P. (1999). Comparison of GENEHUNTER and MFLINK for analysis of COGA linkage data. Genetic Analysis Workshop 11: Analysis of genetic and environmental factors in common diseases. L. Goldin, C. Amos, G. Chase *et al*. *Genet. Epidemiol.* **17**, S115–S120.

Dizier, M. H. & Clerget-Darpoux, F. (1986). Two-disease locus model: sib pair method using information on both HLA and Gm. *Genet. Epidemiol.* **3**, 343–56.

Ertekin-Taner, N., Graff-Radford, N., Younkin, L. H., Eckman, C., Baker, M., Adamson, J., Ronald, J., Blangero, J., Hutton, M. & Younkin, S. G. (2000). Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Science* **290**, 2303–4.

Kehoe, P., Wavrant-De Vrieze, F., Crook, R., Wu, W. S., Holmans, P., Fenton, I., Spurlock, G., Norton, N., Williams, H., Williams, N., Lovestone, S., Perez-Tur, J., Hutton, M., Chartier-Harlin, M. C., Shears, S., Roehl, K., Booth, J., Van Voorst W., Ramic, D., Williams, J., Goate, A., Hardy, J. & Owen, M. J. (1999). A full genome scan for late onset Alzheimer's disease. *Hum. Mol. Genet.* **8**, 237–45.

Lin, S. (2000). Monte Carlo methods for linkage analysis of two-locus disease models. *Ann. Hum. Genet.* **64**, 519–532.

Mein, C. A., Esposito, L., Dunn, M. G., Johnson, G. C.,

Timms, A. E., Goy, J. V., Smith, A. N., Sebag-Montefiore, L., Merriman, M. E., Wilson, A. J., Pritchard, L. E., Cucca, F., Barnett A. H., Bain, S. C. & Todd, J. A. (1998). A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nat. Genet.* **19**, 297–300.

Myers, A., Holmans, P. & Marshall, H. *et al.* (2000). Susceptibility locus for Alzheimer's disease on chromosome 10. *Science* **290**, 2304–5.

Ott, J. (1991). *Analysis of Human Genetic Linkage* (revised edition). Baltimore, Johns Hopkins University Press.

Risch, N. (1989). Linkage detection tests under heterogeneity. *Genet. Epidemiol.* **6**, 473–80.

Schork, N. J., Boehnke, M., Terwilliger, J. D. & Ott, J. (1993). Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. *Am. J. Hum. Genet.* **53**, 1127–36.

Sham, P., Zhao, J. & Curtis, D. (2000). Power comparison of parametric and non-parametric linkage tests in small pedigrees. *Am. J. Jum. Genet.* **66**, 1661–8.

Smyth, C., Kalsi, G., Curtis, D., Brynjolfsson, J., O'Neill, J., Rifkin, L., Moloney, E., Murphy, P., Petursson, H. & Gurling, H. (1997). Two-locus admixture linkage analysis of bipolar and unipolar affective disorder supports the presence of susceptibility loci on chromosomes 11p15 and 21q22. *Genomics* **39**, 271–8.

Tienari, P. J., Terwilliger, J. D., Ott, J., Palo, J. & Peltonen, L. (1994). Two-locus linkage analysis in multiple sclerosis (MS). *Genomics* **19**, 320–5.