Published in final edited form as:

Genet Epidemiol. 2008 December ; 32(8): 800–815. doi:10.1002/gepi.20350.

Multipoint Lods Provide Reliable Linkage Evidence Despite Unknown Limiting Distribution: Type I Error Probabilities Decrease with Sample Size for Multipoint Lods and Mods

Susan E. Hodge^{1,2,3,*}, Laura Rodriguez-Murillo³, Lisa J. Strug⁴, and David A. Greenberg^{1,2,3}

¹ Division of Epidemiology, NY State Psychiatric Institute, New York

² Department of Psychiatry, College of Physicians & Surgeons, Columbia University, New York

³ Division of Statistical Genetics, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York

⁴ The Hospital for Sick Children, University of Toronto, Toronto

Abstract

We investigate the behavior of type I error rates in model-based multipoint (MP) linkage analysis, as a function of sample size (N). We consider both MP lods (i.e., MP linkage analysis that uses the correct genetic model) and MP mods (maximizing MP lods over 18 dominant and recessive models). Following Xing & Elston [2006], we first consider MP linkage analysis limited to a single position; then we enlarge the scope and maximize the lods and mods over a span of positions. In all situations we examined, *type I error rates decrease with increasing sample size*, apparently approaching zero. We show: (a) For MP lods analyzed only at a single position, well-known statistical theory predicts that type I error rates approach zero. (b) For MP lods and mods maximized over position, this result has a different explanation, related to the fact that one maximizes the scores over only a finite portion of the parameter range.

The implications of these findings may be far-reaching: Although it is widely accepted that fixed nominal critical values for MP lods and mods are not known, this study shows that whatever the nominal error rates are, the actual error rates appear to decrease with increasing sample size. Moreover, the actual (observed) type I error rate may be quite small for any given study. We conclude that multipoint lod and mod scores provide reliable linkage evidence for complex diseases, despite the unknown limiting distributions of these multipoint scores.

1. Introduction

Although a great deal is known about the statistical behavior of both lods and mods in two-point linkage analysis [e.g., Ott, 1999; Hodge et al., 1997; Strauch, 2007], little has been investigated about either lods or mods for multipoint (MP) linkage analysis. Greenberg & Abreu [2001] investigated the precision with which MP mods could determine trait locus position. (They also examined error rates, but only in a limited way.) A recent paper by Xing & Elston, 2006 (henceforth X&E) attempts to fill this gap. These authors examined the null distribution of MP lod scores, when those lods are evaluated at a single marker location (i.e., are not maximized over position). This distinction of evaluating lods or mods at a single point vs. maximizing them over position is critical, because in practice, investigators do not evaluate MP lod scores at only one position. Thus, it is unclear how much one can learn

*Correspondence, E-mail: seh2@columbia.edu.

about the type I errors associated with MP analysis, from the X&E study. In this paper, we examine the behavior of type I error for position-maximized lod and mods as well as single-position ones. (We use “single-position” to refer to MP scores that are evaluated at only one position, and “position-maximized” for MP scores that are maximized over position, whether over a small or large region.)

X&E performed both theoretical analysis and simulations. In the analytic part of their study, they pointed out that the null distribution of single-position lods is approximately normally distributed, with mean and variance that depend on factors specific to the disease model and study design (e.g. sample size, genetic parameters, etc.) Via computer simulation they studied the behavior of type I errors for MP single-position mod scores. They concluded that when false positives occur, they tend to appear more frequently for some analysis models than others. However, we find it curious that they focused on the model under which the false positive occurred, and not simply on the fact of type I error. Moreover, we note that their conclusions are limited to the single-position situation and to their limited set of simulations. These conclusions cannot necessarily be generalized to position-maximized MP linkage analysis.

In this paper we extend the X&E study, with two aims: (1) For *single-position* lods, to provide a more general analytic result, one that comes from standard statistical theory and that explains the general behavior of these lods, and for mods, to extend the simulations; (2) For *position-maximized lods and mods*, to expand X&E’s simulations, in order to evaluate the statistical behavior of these scores when they are maximized over position and to evaluate what this implies for their use.

Finally, X&E focused on the question of which analysis model is “chosen” by a mod-score analysis, *given* that the analysis has yielded a significant false positive score. Although our focus was on the magnitude of type I error, we examined this question as well.

We conclude that even though the limiting distributions of MP position-maximized lods and mods are unknown, not only do these MP scores provide reliable linkage evidence, but also type I error rates decrease with increasing sample size.

2. MP Linkage Analysis at a Single Position

2.1. Theoretical Underpinnings

2.1.1. Notation—Let $MPL(\xi)$ denote the multipoint likelihood evaluated at ξ , where ξ represents position along a chromosome. We arbitrarily specify the leftmost marker to be at position 0; then ξ gives the distance of any point from that marker, either to the right (+) or to the left (−), as measured in map units (centiMorgans, cM, or Morgans, M). The investigator’s ultimate goals are to determine the most likely value of ξ for the trait locus and to evaluate the strength of the evidence at that position.

We can represent the null hypothesis of no linkage as $MPL(\pm\infty)$ [Ott, 1999] Then the multipoint lod score, $MPLod(\xi)$, is defined as this log-likelihood ratio (log-LR):

$$MPLod(\xi) \equiv \log [MPL(\xi)/MPL(\pm\infty)]. \quad (1)$$

(Throughout this paper, “log” indicates common logs, base 10.) For the mod score, we explicitly incorporate the genetic model into the likelihood, writing the likelihood as $L(\xi, \phi)$. Here ϕ indicates all other genetic parameters besides the position ξ (mode of inheritance,

penetrances, multiple loci, etc.) Then the the multipoint mod score, $\text{MPMod}(\xi)$, represents the MP lod maximized over the genetic models:

$$\text{MPMod}(\xi) \equiv \max_{\phi} \log [\text{MPL}(\xi, \phi) / \text{MPL}(\pm \infty, \phi)]. \quad (2)$$

We do not perform a true maximization over all possible models; rather, we maximize over nine dominant and nine recessive models [cf. Hodge et al., 1997; Xing & Elston, 2006].

2.1.2. Behavior of type I error for multipoint lod scores calculated at a single position—The log-LR in eq. (1) represents a ratio of likelihoods for *two simple hypotheses*, as mentioned by Xing & Elston [2006]. While the “usual asymptotic theory for likelihood ratio tests does not hold,” [X&E, p 449, left column] we do know that the probability of making a type I error for this kind of LR decreases as sample size (N) increases, i.e., as information increases [e.g., Royall, 1997; Strug & Hodge, 2006]. Moreover, for any desired cutoff value $k > 1$ and for any specified $\varepsilon > 0$, one can make N large enough to ensure a probability greater than $1 - \varepsilon$ that a sample of size N will produce a LR of at least k in favor of the true hypothesis [Royall, 2000]. Thus, the probability that the MP lod in eq. (1) will exceed any given cutoff value, when there is no linkage, must approach zero as N increases. Elsewhere, we have illustrated this phenomenon for two-point lod scores [Strug & Hodge, 2006]. In Sec. 2.2 we will present results of computer simulations to demonstrate this behavior and to assess how quickly the error rate approaches zero in MP analysis.

In particular, the fact that the distribution of the log LR derived by X&E is a function of sample size and of other genetic parameters is consistent with Royall’s [2000] results for quite general likelihoods. As a consequence, the distribution for the type I error rate is also characterized, and the maximum value of this distribution is bounded and converges to zero with increasing sample size/information. Therefore, differential error rates under the null can be explained by differences in the quantity of *information*, whether those differences are reflected in different sample sizes or in more- or less-informative genetic models. X&E observed that the error rate varies with sample size but did not note the more critical fact of convergence to zero as sample size increases.

2.1.3. Behavior of type I error for multipoint mod scores calculated at a single position—The theoretical behavior of MP mods is more complicated. The quantity in eq. (2) is a conditional profile likelihood, and as such, existing statistical theory does not ensure that the error probability will approach zero as N increases. Nevertheless, it seems reasonable to expect that error probabilities for mods, though larger in absolute magnitude than for lods, would also approach zero with increasing sample size, as the lod error probabilities do. We investigate this behavior via computer simulations in Sec. 2.2.

2.2. Computer Simulations

2.2.1. Assumptions and methods—*Markers*: Except where noted otherwise, we simulated two marker loci, A and B , separated by a distance of ≈ 10 cM. The marker loci are in linkage equilibrium. *Trait*: We generated data using a number of different trait models, each referred to as a “true” or “generating” model (GM). The trait was due to a gene showing either a dominant or a recessive mode of inheritance, with reduced penetrance, and the trait locus was not linked to either marker. Let T represent the trait allele and t , the wild type. For the dominant generating models (GMs), the frequency of T was 0.01, and the penetrance vectors for genotypes (TT , Tt , tt) were (f , f , 0.001), where f assumed the values 0.20, 0.50, or 0.80 (models D20, D50, and D80, respectively). For the recessive GMs, frequency of T was 0.14, and penetrance vectors were (f , 0.001, 0.001), where f assumed the same three values (models R20, R50, and R80, respectively). *Families*: We simulated four-

child families with at least two affected children. Parents were typed and were fully informative at both marker loci. That is, in every nuclear family, the father is “a₁a₂” at the *A* marker locus, and “b₁b₂” at the *B* locus, whereas the mother is “a₃a₄” at *A* and “b₃b₄” at *B*. Parental affectedness status was known. *Datasets*: Dataset sizes varied, with each dataset containing *N* families, *N* = 10, 20, 30, 40, 50, 75.

We analyzed the simulated data assuming a range of analysis models (AMs), among which the true model (GM) was always included. Analysis Models D10, D20, ..., D90 had penetrance vectors (*f*, *f*, 0.001) where now *f* assumed values 0.1, 0.2, ..., 0.9; and similarly for AMs R10 through R90. Gene frequency was always the same as in the respective generating model. (We did not consider higher gene frequencies because of linkage analysis's low power when gene frequency is high and penetrance is low.) For each combination of GM and dataset size *N*, we simulated 1000 datasets. For each dataset, first the MP lod score was calculated, at the specified position, with the assumed AM set equal to the GM. Then each dataset was analyzed via MP mods; i.e., we determined the lod score for each of the 18 different AMs, then determined the maximum of those 18 lod scores. For these single-position analyses, scores were calculated at only one position, located halfway between the two markers.

To determine type I error for each combination of GM and dataset size (*N*), we tabulated what proportion of datasets yielded a lod or mod greater than the specified cutoff value. We used cutoff values of 0.9 and 1.5, corresponding to LR_s of ≈ 8 and ≈ 32 , respectively. (If the test statistic, appropriately transformed, followed χ^2 (1 df) distribution, these cutoff values would correspond to one-sided test sizes α of ≈ 0.02 and ≈ 0.004 , respectively.)

2.2.2. Simulation results for single-position analyses—Table 1 shows the results. In part (a), for lods, we report results only for *N* = 10, 20, and 30, and only for the 0.9 cutoff value, since all type I error probabilities quickly went to zero. Part (b) shows the mod results, for *N* up to 75, for both cutoff values. The pattern is less clearcut than with the lods. Nevertheless, within the limits of simulations, the type I error probabilities decrease with increasing sample size for mods, just as they did for the lods, and they also, like the lod error probabilities, appear to approach zero as *N* increases. Figure 1 presents these results (for the 0.9 cutoff) in graphic form.

3. MP Linkage Analysis Maximized Over Position

3.1. Theoretical Underpinnings

3.1.1. Notation—In this section we consider the usual situation, in which investigators maximize MP lod or mod scores over position ξ – either in a genomewide linkage scan, or over a limited region where they suspect their locus of interest is located. Let *G* denote the region over which the scores are maximized. In this context, ξ is our parameter of interest, and we maximize MPLod(ξ) and MPMod(ξ) over ξ in *G*:¹

$$\max \text{MPLod}(\xi) \equiv \max_{\xi \in G} \log [\text{MPL}(\xi)/\text{MPL}(\pm\infty)] \quad (3)$$

¹Ordinarily, mods are described as follows: First one maximizes one's lods over “position” (or over recombination fraction in two-point linkage analyses), for a given genetic analysis model; this yields a “max lod” for that model. Then one chooses the maximum of these maxlods, and this maximum becomes the “mod” or “MMLS” score. In contrast, here we conceptualize the mods as if one first maximizes over model, and then maximizes over position, i.e., reversing the order of the maximizations. Interestingly, this is actually how Hodge & Elston [1994] defined mod scores (page 332). However, the results are the same either way. The reason we refer to them as “maximizing mods over position” is to draw the contrast with Xing & Elston's study, where the authors considered only a single position.

$$\max \text{MPMod}(\xi) \equiv \max_{\xi \in G, \phi} \log [\text{MPL}(\xi, \phi) / \text{MPL}(\pm\infty, \phi)]. \quad (4)$$

3.1.2. The MPLod(ξ) function—Theoretically, the parameter space for ξ is defined as $-\infty < \xi < \infty$, but for MP linkage analysis, this space is truncated, depending on the region under investigation. The Appendix gives the functional form of MPLod(ξ) when there are only two marker loci (eq. A-9), and Figure 2 graphs several possible manifestations of the MP lod score for the two-marker case. (We discuss the figure in more detail below, in Sec. 3.3.3, Example 2). Note that the function can be multimodal and can have discontinuities, where the value of the function approaches $-\infty$. Most importantly, the null hypothesis corresponds to ξ approaching $-\infty$ or $+\infty$. Thus, in contrast to two-point linkage analysis, where the null hypothesis corresponds to $\theta_0 = 0.5$, for MP linkage analysis there is no single, well-defined null-hypothesis “value” for ξ .

From the mathematical statistical viewpoint, $\hat{\xi}$, the MLE (maximum likelihood estimator) of ξ , is not guaranteed to be consistent (asymptotically unbiased) on a truncated parameter space, and the log-L function violates several of the usual regularity conditions for asymptotic normality of the MLE. Therefore, the usual result for a log-LR with parameter θ —that under the null hypothesis, $2\ln[L(\hat{\theta})/L(\theta_0)]$ asymptotically approaches χ^2 distribution with appropriate degrees of freedom—does not hold [Self & Liang, 1987]. Furthermore, it seems reasonable that, when the null is true, the value of $\hat{\xi}$ in any one dataset may often occur outside our range of evaluation, i.e., outside G . This is because one cannot evaluate the log-LR over ξ ’s entire parameter space from $-\infty$ to $+\infty$. Because of these issues, we decided to investigate the behavior of type I error probabilities for position-maximized MP lods and mods via simulation.

Note that from the biological viewpoint, there is nothing unusual about the null hypothesis. It is simple and well-characterized, and one can straightforwardly calculate the probability of observations under the null (as in eq. A-2 in the Appendix). It is only in terms of its statistical properties that the null hypothesis is problematic.

It is clear from these considerations that single-position and position-maximized lods and mods have different statistical properties. As we have discussed above, and as X&E also point out, position-maximized lods and mods do not have a known limiting distribution, in contrast to the single-position scores. Therefore, we now investigate the behavior of position-maximized lods and mods separately.

3.2. Computer Simulations

3.2.1. Assumptions and methods—GMs and AMs were the same as in Sec. 2. We maximized lods and mods over a grid of 51 equispaced points on the 10 cM span between marker loci A and B , and we evaluated type I error rates, for dataset sizes up to $N = 75$.

We also performed a few lod analyses over wider ranges, and with more markers, in order to see if the results differed substantially. We did these for just two GMs: D50 and R50. For these simulations, we generated 18 equispaced markers spanning 50 cM (evaluated at 86 points), and also 50 markers spanning 100 cM (evaluated at 246 points), then calculated max MP lod scores and type I error rates for dataset sizes up to $N = 50$. All markers are in linkage equilibrium.

3.2.2. Simulation results for position-maximized analyses—Table 2(a) shows the lod results for the 10 cM span. We report results only up to $N = 40$, since the error probabilities quickly went to zero. Part (b) shows the mod results for the same span. Again,

type I error rates keep dropping as sample size increases. For the lod, the drops in error rates are steep and monotonic. For the mods, the results are less definitive than for the lods but the trend is clear. Error rates are higher for models with low penetrance (except for the mods for recessive data), but in all cases, these rates drop with increasing N . Figure 3 presents the results for the 0.9 cutoff in graphic form.

Maximizing lods over larger spans and with more markers did not change the pattern of the results (Table 3 and Figure 4). Error rates were high for small N (10 families) but dropped off faster than for the narrower spans and also appear to approach zero rather quickly. As an interesting side point, these results also imply that, in small samples, shorter chromosomes may have lower error rates than longer ones.

3.3. Explanation of the Findings

Our results for the position-maximized analyses appear less straightforward than those for the single-position analyses. We originally expected a maximized LR to have some kind of “stable” asymptotic type I error rate, one that does not vary with sample size, once some reasonable sample size has been reached. Instead, we observed error rates that appear to approach zero with increasing sample size.

We believe that this behavior is explained by the fact that the null hypothesis corresponds to ξ approaching $\pm\infty$, as opposed to equaling some fixed finite value in the region of evaluation (G). To understand this, we distinguish between the “theoretical” or “nominal” maxlod (i.e., the maximum over the whole range of ξ , from $-\infty$ to $+\infty$), and the “apparent” or “observed” maxlod (i.e., the maximum only over the range of ξ that one is examining, whether that range G consists of 10 cM or 100 cM or more). The theoretical maxlod represents a global maximum, whereas the observed maxlod might be only a local maximum. Similarly, we distinguish between the corresponding “theoretical” and “observed” values of $\hat{\xi}$, the MLE of ξ .

We hypothesize that under the null hypothesis, as N increases, the theoretical $\hat{\xi}$ moves closer to $\rightarrow\pm\infty$. In other words, as N increases, $\hat{\xi}$ moves farther to the right or to the left (or sometimes equally to both, when the data are symmetric²), and thus eventually $\hat{\xi}$ moves *outside the region, G , that one is examining* (no matter how large G is). We further hypothesize that the probability of $\hat{\xi}$ falling outside any given region G increases as sample size increases, and that this is also true if we consider only outcomes for which the max lod exceeds the cutoff value. That is, we speculate that $\Pr[(\text{maxlod} \geq \text{cutoff}) \cap (\hat{\xi} \text{ outside } G)]$ increases with increasing N , for any specified G .

In other words, even if there does exist some “stable” asymptotic test size α associated with the maximum MP lod score, then

$$\alpha \equiv \Pr[\text{maxlod} \geq \text{cutoff}]$$

$$\alpha = \Pr[(\text{maxlod} \geq \text{cutoff}) \cap (\hat{\xi} \text{ inside } G)] + \Pr[(\text{maxlod} \geq \text{cutoff}) \cap (\hat{\xi} \text{ outside } G)] \quad (5)$$

If the second probability on the right side of (5) increases with increasing N , as we hypothesize, this means that the first probability on the right side of (5) must decrease with increasing N . And that first probability is our apparent (observed) type I error rate.

²By “symmetric data” we mean datasets in which $n_{10} = n_{01}$, i.e., the number of meioses representing recombination with locus A and not with locus B equals the number representing recombination with B and not with A . In these cases, the MP curve to the right of B exactly mirrors that to the left of A .

In short, according to this explanation, the theoretical type I error rate may well approach some stable nominal value as N increases, but the observed error rate, given that one observes only a finite region G , decreases as the theoretical ξ moves farther away from G . We cannot call on consistency of the MLE to justify this explanation [Self & Liang, 1987]. Nevertheless, our explanation is supported both by common sense and by the behavior of the two special cases that we now examine.

3.4. Two special cases

To assess the feasibility of this explanation, we examined two examples. One involves two-point lod scores, and the other involves a multipoint analysis with two marker loci.

Example 1: Two-point lod scores—Consider standard two-point lods, maximized over recombination fraction θ , and with the lod score calculated over the full range of $\theta \in [0.0, 0.5]$. What would happen to one's type I error rate if one examined these two-point lod scores only in the range $G = [0.0, 0.2]$? By the same reasoning as above, the "observed" type I error probability in that case should decrease with increasing N , because again, one would be looking at error rates based on a local rather than global maximum. Under the null hypothesis, the larger the sample size, the higher the probability that the global maximum will occur outside G , since $\hat{\theta}$ is consistent and the true (null-hypothesis) value, $\theta = 0.5$, lies outside G .

To confirm this prediction, we simulated nuclear families of mixed sizes, under fully penetrant dominant inheritance; parents were fully informative at the marker locus and had known affectedness status. (The family-size distribution is taken from Cavalli-Sforza & Bodmer, 1971, with a mean family size of 2.8 and standard deviation of 2.3.) For each value of N ($N = 10, 20, \dots, 50$), we simulated 2,000 datasets of nuclear families. We tabulated the behavior of type I error rates with increasing sample size, for cutoff values of 0.59, 0.90, and 1.17. Table 4 shows the results, confirming that the error rates do decrease with sample size. This lends support to our explanation. (Interestingly, the observed type I error rates are relatively close to their nominal values for datasets of size $N = 10$ but are already considerably smaller by the time N reaches 20.) In contrast, if we had maximized the lods over the entire range $[0.0, 0.5]$, the observed type I error rates would have adhered very closely to their nominal values, as is well known.

Example 2: Multipoint linkage analysis with two marker loci—The two-point example confirms that our explanation is reasonable in principle. We can also go further and analytically investigate the exact behavior of the *multipoint* lod score function over its entire range when there are only two marker loci (A and B) and they are separated by a reasonably small (≤ 10 cM) distance d . We assume Haldane's map function (i.e., no interference) [Sham, 1998]. Without loss of generality, we specify that A is located at $\xi = 0$ and B , at $\xi = d$. We distinguish three regions: " L " (to the left of Locus A , i.e., $\xi < 0$), " W " (between the two loci, $0 < \xi < d$), and " R " (to the right of B , $\xi > d$). Except in certain special circumstances (e.g., Fig. 2e, f), the MP lod score is undefined ($\rightarrow -\infty$) at $\xi = 0$ and $\xi = d$.

Figure 2 illustrates several different possible multipoint lod score curves and the datasets that can yield them, with $d = 0.1$ M. Let n_{11} represent the number of meioses resulting from a recombination of the trait locus T with both loci; n_{10} , recombination with locus A only; n_{01} , recombination with B only; and n_{00} , no recombination with either locus. Note $N = n_{11} + n_{10} + n_{01} + n_{00}$.

- The first four examples (Fig. 2a–d) illustrate situations in which there is at least one recombination between T and each marker locus, i.e., where both $n_{11} + n_{10} > 0$ and $n_{11} + n_{01} > 0$. There are discontinuities at $\xi = 0$ and $\xi = d$, with the function

approaching $-\infty$ from both sides. The function in region L can either (i) approach a maximum of 0 from below, as ξ approaches $-\infty$ (Fig. 2a); or (ii) have a positive maximum at some value of $\xi < 0$, then approach 0 from above as ξ approaches $-\infty$ (Fig. 2 b, c, d). Similarly, in region R , the function can either (i) approach a maximum of 0 from below, as ξ approaches $+\infty$ (Fig. 2a, b) or (ii) have a positive maximum at some value of $\xi > d$, then approach 0 from above as ξ approaches $+\infty$ (Fig. 2 c, d). In region W , the function always has (at least) one local maximum for some value(s) of ξ , and the lod score may be either negative (Fig. 2a, c) or positive (Fig. 2b, d) at this maximum.

- The last two examples (Fig. 2e–f) illustrate situations in which one or both marker loci do not recombine with T . Fig. 2e shows a possible outcome when T recombines with the B locus but not with A ; Fig. 2f shows the shape of the multipoint lod curve when there are no recombinations with either marker locus.

These examples illustrate that up to three distinct maxima can exist, one each in regions L , W , and R . Our question is: How often does the maximum that one *observes* – the one in W , the region between the markers – coincide with the global maximum (as opposed to the global maximum occurring in L and/or R)? More specifically, how often does this happen when the global maximum is greater than or equal to our cutoff value for significance?

We have analytically derived the probabilities (under the null hypothesis of no linkage) for the global maximum to be greater than any specified cutoff value, and to occur inside W , as opposed to outside W . These are exact calculations of probabilities, not simulations; details are in the Appendix. The probabilities are functions of the specified cutoff value, of d (the distance between the markers A and B), and of the sample size N . Table 5 shows these probabilities as functions of N , for a cutoff value 0.9, for $d = 0.10, 0.05$, and 0.01 . In the table, the column headings refer to, respectively, the following seven events: “ L ” = $\{(\text{global max lod} \geq 0.9) \cap (\text{global max lod is in region } L)\}$; similarly for “ W ” and “ R ”; “ LW ” = $\{(\text{global max} \geq 0.9) \cap (L \text{ and } W \text{ are tied for the global max})\}$; similarly for “ LR ” and “ WR ”; and $LWR = \{(\text{global max} \geq 0.9) \cap (\text{all three regions are tied})\}$. The entries in the table give the respective probabilities of each event. For example, consider the row for $N = 10$ in part (a) of the table: If one looked at ξ only between the two marker loci (region W), one’s apparent or observed type I error rate would equal the sum of the values in the four **bold** columns, i.e., $0.0007(W) + 0.0006(LW) + 0.0006(WR) + 0.0004(LWR) = 0.0023$. The remaining 0.0140 of the error rate occurred in regions L (0.0051), R (0.0051), or “tied” in L and R (0.0038), where one would not “see” it. These results are exactly analogous to those from the two-point example (Example 1), in which we ignored part of the possible range of the recombination fraction θ .

Table 5 confirms that for each parameter set considered, the observed type I error rate, represented by the sum of the probabilities in the W , LW , WR , and LWR columns, quickly goes to zero as N increases. This observed error rate has reached zero by $N = 20$ when $d = 0.10$, and reaches zero even more quickly for smaller values of d . (The observed error rates are even smaller for a cutoff value of 1.5; not reported in the table.)

We speculate that the same principle explains the mod score results as well: As sample size (or information) increases, the maximum mod falls outside the region being analyzed (G) with higher probability, and hence the “observed” type I error rate decreases. It is true that in any given dataset, type I error is larger (on average) for a mod analysis than for a lod analysis, since one is maximizing over multiple models in the mod analysis, but the downward trend is similar.

Interestingly, from these data, we can also investigate the behavior of the total (i.e., “theoretical”) type I error rate, when the behavior of ξ is evaluated over its entire range. For any given row in Table 5, the sum of the seven probabilities in that row yields $\Pr[\text{global max lod} \geq \text{cutoff value}]$. Table 6 shows these sums as a function of N , up to $N = 100$, for different marker distances d and for two cutoff values, 0.9 and 1.5. Again, these are exact calculated error rates, not simulated. We considered two marker loci separated by distances of $d = 10, 5, 1, 0.1$, and 0.01 cM, and for sample sizes of 10, 20, ..., 100 informative meioses.

These total error rates in Table 6 are rather close to what one would expect if the test statistic followed a one-sided χ^2 distribution with one degree of freedom. Consider the 0.9 cutoff value (Table 6a), for which the χ^2 type I error rate would be ≈ 0.02 . When the two marker loci are separated by 10 cM, our calculated error rates, though somewhat larger than 0.02, are in the approximate range (between .025 and .035 for $N \geq 70$). When the two markers are closer together ($0.10 - 0.01$ cM), our calculated error rates remain between 0.017 and 0.023 for $N \geq 70$. Table 6b reveals a similar pattern for the 1.5 cutoff value, for which the corresponding χ^2 -based error rate would be ≈ 0.004 . Our calculated rates are between 0.005 and 0.009 when $d = 10$ cM, and between 0.004 and 0.006 for small d (for the larger dataset sizes).

It is not surprising that the error rates would be close to the χ^2 error rates for two marker loci that are very close together (i.e., $d \rightarrow 0$), since then the two loci function essentially as a single locus, and the MP lod score approaches a two-point lod (see Sec. A.4, Appendix). What is interesting is how, even for intermarker distances up to $d = 10$ cM, these type I error levels are still relatively close to those associated with the one-sided χ^2 with 1 d.f., though slightly larger.

4. Comparison with Xing & Elston [2006]

As mentioned in the Introduction, our focus is different from that of X&E. They looked primarily at how often a false positive mod score pointed toward the wrong genetic model, whereas we have focused on the magnitude of false positive rates and how these error rates behave with increasing sample size. Nevertheless, since both studies reach conclusions regarding the utility of multipoint mod analysis, they should be compared.

4.1 Single-position vs. position-maximized lod scores

The stated purpose in X&E is, “to study the distribution of multipoint lod scores both analytically and by simulation,” and also to “study by simulation the distribution of maximum multipoint lod scores when maximized over different penetrance models” [Abstract]. In their paper, the authors perform most of their analyses on single-position lods and mods. Therefore, their results cannot be assumed to apply to multipoint analyses maximized over position.

The authors state (page 449, left) that, “... once θ , the recombination fraction between the trait locus and one specific marker, is fixed, given a particular map function the relationships between the trait locus and all markers in the map are fixed; therefore, we can simply focus on only one marker and a trait locus to study the properties of the multipoint lod score.” They continue, “In this setup H_1 is also a simple hypothesis and the usual asymptotic theory for likelihood ratio tests does not hold.” However, we maintain that focusing on only one marker locus is a separate issue from performing a *single-position* linkage analysis. (After all, in classical two-point lod analysis, which also involves only one marker locus, one still maximizes one’s lods over position, where position is represented by recombination fraction $\theta \in [0.0, 0.5]$.) As noted above, results from single-position linkage analysis would not apply to position-maximized linkage analysis, either for two-point or MP analysis.

In this paper we have examined the behavior of both single-position and position-maximized lod scores and mods. We considered the first case (Sec. 2) so that we could replicate and extend X&E's findings. However, investigators rarely use lod scores or mods that way, so we devoted most of this study to the second case (Sec. 3). (Conceivably, one might limit one's MP lod or mod calculations to a single position in the case of a candidate gene, but this is not the usual way to perform MP linkage analysis.)

4.2. False positives vs. wrong models

X&E examined the conditional distribution of max MPMOD scores given that these scores equal or exceed a given cutoff value (i.e., the distribution of statistically significant false positives), in order to determine how often the max mod score is associated with the wrong generating model. For simplicity, we use the term "estimated model" to denote the analysis model associated with the max mod score, when the mod has been maximized over analysis models.

For example, in their Table IV, when their generating model was D50, they observed that 41 out of 1,000 simulated 20-family datasets yielded a max mod above the cutoff value of 0.83 (i.e., the observed type I error rate for that cutoff value was 0.041). Only 17 of these 41 "false-positive" datasets resulted in any of the dominant models being the estimated model; the remaining 25 yielded a recessive estimated model. Moreover, of those 17 estimated models that were dominant, 12 had a penetrance of 30% or lower. Thus, in that example, the estimated model was "incorrect" over half the time, and even when it was correct, the penetrance was too low. X&E reported a general pattern in which the estimated model tended most often to be either dominant with low penetrance or recessive with high penetrance, regardless of generating model.

We examined the same conditional distributions of max mods and estimated models, but under broader conditions. Table 7(a) shows our results for the single-position mods, for 20-family datasets. The table follows the same format as X&E's Table IV, showing the conditional distribution of MPMODs, conditional on attaining pre-specified cutoff values, classified by their corresponding estimated models (pooled three at a time). Thus, Table 7(a) corresponds roughly to X&E's Table IV, with the following differences:

1. We examined only two thresholds, 0.9 and 1.5, corresponding approximately but not exactly to two of their cutoffs, 0.83 and 1.44.
2. We simulated data using a lower "sporadic" rate, 0.001 instead of their 0.01. (They stated that their parameters "mimicked those in Hodge et al. [1997]," but that is incorrect, as the simulations in Hodge et al. used a sporadic rate of zero.) Also, for the recessive GMs, we used a higher frequency for the disease-predisposing allele (0.14 instead of 0.01). We did this to reduce the proportion of affected individuals with the sporadic form of the disease. Particularly in X&E's recessive GMs, with their low gene frequency and high sporadic rate, only a tiny fraction (<1%) of affected individuals have the genetic form of the disease.
3. We included parental affectedness status in our linkage analyses, whereas X&E set parental status as unknown.

In order to see the effect of sample size on the results, we also examined datasets of 40 families each (Table 7b).

Next we extended the investigations from single-position mods to position-maximized mods. We maximized the MP mods over a range of 10 cM that included 11 equispaced marker loci. Table 8 shows these results, for 20-family and 40-family datasets.

In Tables 7 and 8 we observed different patterns of estimated models than those reported by X&E. Low-penetrance AMs are consistently favored much more often than high-penetrance ones, particular as dataset size increases from 20 to 40 families. This does not accord with X&E's observation of recessive models with high penetrance, along with dominant models with low penetrance. Additionally, in all our simulations, for the dominant GMs, error rates decreased as the generating penetrance increased (Tables 7 and 8, Figures 1 and 3), whereas in X&E's simulations, that was not true. These differences may be due to X&E's high sporadic rates, which probably add a good deal of "noise" to the system.

The pattern X&E reported was based on simulations performed under a narrow set of conditions and apparently does not hold in general. Also, some of the patterns in both studies may depend on the fact that only 4-child families with at least two affected children were examined.

5. Discussion

5.1. Multipoint Lods and Mods Evaluated at a Single Position

We began (Sec. 2) by examining the behavior of MP lods and mods evaluated at a fixed position, i.e., single-position linkage analysis. For the lods, this involves a likelihood ratio of two *simple* hypotheses, and therefore the error rate necessarily approaches zero as sample size increases [Royall, 1997; Strug & Hodge, 2006]. Our simulations show that this approach to zero occurs rapidly as a function of sample size. For the mods, the error rates also appear to approach zero with increasing sample size, but not as steeply as for the lods.

The likelihood ratio (LR) for two simple hypotheses has different statistical properties than the more commonly computed LR that compares the likelihood at the MLE with the likelihood at the null-hypothesized value of the parameter. It is the LR for two simple hypotheses that X&E concentrate on in their paper, providing an analytical result and simulations to illustrate its behavior. The statistical properties and behavior of this LR are already well-known in the statistical literature [e.g., Smith, 1953; Birnbaum, 1962; Hacking, 1965; Royall, 2000; Blume, 2002]. Moreover, its use in the analysis of genetic data [Vieland & Hodge, 1998] and specifically in two-point linkage analysis [Strug & Hodge, 2006] has been delineated. However, as we discuss in the present paper, the LR for two simple hypotheses has limited applicability in multipoint linkage analysis, where investigators examine their MP lods and mods evaluated at different positions and choose the maximum among those.

5.2. Multipoint Lods and Mods Maximized Over Position

Our position-maximized multipoint simulations also showed that type I error decreases as sample size increases. This decrease is striking for the max MPLods, and it is also clear for the max MPMods (though less pronounced). We were surprised by this phenomenon at first.

Clearly, whatever this phenomenon was that we were observing, it was unrelated to the phenomenon described above for two *simple* hypotheses (Secs. 2.1 and 5.1). We hypothesized that under the null hypothesis of no linkage, $\hat{\xi}$ falls farther outside our region G as sample size goes up, moreover, that the probability of falling outside G increases with increasing sample size. Thus, whereas there may indeed exist a total "stable" asymptotic test size α for the MP lod score, the "apparent" or "observed" error rate associated with any dataset must be less than α and, moreover, appears to approach zero as sample size increases. In Sec. 3.3, we explained this in more detail and in Sec. 3.4 we illustrated how our explanation appears to account for our observations.

We note that our second example (the multipoint example) in Sec. 3.4 used the Haldane map function, which assumes no interference (see Appendix). For the region G between the two marker loci, the choice of map function has a negligible effect, since the markers are reasonably close together (≤ 10 cM). For the regions outside G , other map functions (e.g., Kosambi, Sturt, etc. [Sham, 1998]) will give somewhat different numbers than those we report here. However, the point of the example is to illustrate the principle that ξ moves farther outside G as sample size increases, and that principle will hold, whatever map function one uses.

We were also curious as to whether a “stable” test size exists for MP lods maximized over position, when the lods are maximized over their entire range ($-\infty$ to $+\infty$). Do error rates approach any constant value with increasing sample size? In Table 6 we tabulated the “total” (i.e., theoretical) type I error levels – based on exact calculations, not simulations – for lods in a multipoint analysis involving two marker loci. The resultant error levels were fairly close to 0.02 for the 0.9 threshold, and to 0.004 for the 1.5 threshold, which are, respectively, the error rates associated with the one-sided χ^2 (1 d.f.). The situation examined in Table 6 is too limited to generalize to all multipoint lod analyses, but it provides intriguing evidence that perhaps the *theoretical* type I error rates do exhibit stable behavior in the limit.

One more comment about maximizing over position: Investigators often think of MP linkage analysis in terms of each individual lod score, as it is calculated at a single position. This can be a useful way to think, for example in the context of misspecification of genetic parameters, since it reminds the investigator that one does not have recombination fractions to compensate for misspecification [e.g., Risch & Giuffra, 1992; X&E, page 454, right column]. However, when it comes to determining the *most likely* position of the trait gene being studied, one must view the multipoint lod scores as representing a function of ξ , a function that one then maximizes with respect to ξ . This point may seem obvious to the reader, but we mention it explicitly, since this reasoning provided the impetus for us to express the MP lod function and examine its statistical behavior in Sec. 3.4, Example 2, and in the Appendix.

5.3. Implications and Applications

Investigators may feel frustrated by not having a known limiting probability distribution to refer to when they perform multipoint linkage analyses, in order to obtain a p -value. This is particularly true when these are “mod,” not “lod” analyses, since the statistical properties of mods are not clear, even in two-point linkage analysis. The surprising result we have shown here is that even when one does not know the absolute type I error rate, one can apparently be confident that the error rate will decrease with increasing sample size/information.

Readers may also be surprised at just how low the error rates were in our simulations – even for position-maximized mod analyses, even with data from low-penetrance genetic models, and even with relatively small datasets. For example, for datasets of only 40–50 families, and data generated by the low-penetrance R20 model, the error rates remained below 0.04, and they were even lower for D20-generated data (below 0.02) – and this is for a conservative cutoff value of 0.9 (Figure 3 and Table 2). When we used 1.5 as the cutoff, error rates did not exceed 0.01 for datasets of 50 families or more, for all genetic models that we examined (Table 2). Future work will need to evaluate more-complex models, such as those with multiple loci, heterogeneity, additivity, etc. Also, our work assumed fully informative markers, no errors in the marker map, etc. When these conditions are relaxed, the absolute error rates will no doubt be larger than those we report; however, we anticipate similar patterns of decreasing error rates with increasing sample size.

In conclusion, we agree with X&E that one should exercise caution in using MP lod and mod scores – because one should exercise caution in *any* analysis of complex diseases with unknown mode of inheritance – but we find their emphasis peculiar. False positives are “bad” already because they are false positives, in this context, giving false indication of linkage. We do not see why they are “worse” because they point to the wrong model in addition to giving false evidence of linkage. Nor is it surprising that when there is no linkage, MP mods do not give reliable indications about mode of inheritance, since – as X&E also point out [page 456, right column] – lod scores do not contain information about mode of inheritance in the absence of linkage [Elston, 1989; Hodge & Elston, 1994]. What we find more interesting and more valuable is how low those absolute type I error rates are and also how these error rates decrease so consistently as sample size increases.

We also disagree with their conclusion that parametric MP analysis should be used only after non-parametric approaches have indicated evidence for linkage. We have demonstrated that MP lod analysis will not lead investigators to conclude linkage with a high probability when, in fact, linkage is not the case. Thus, the MP lod score provides reliable linkage evidence despite its unknown limiting distribution.

Acknowledgments

This research was supported in part by grants: DK-31813 and MH-48858 to SEH, DK-31775 and NS-27941 to DAG, HG00-4314 to LJS, and the Fundacion Alfonso Martin Escudero, Spain (research fellowship to LR-M)

References

- Birnbaum A. On the foundations of statistical inference (with discussion). *J Am Stat Assoc.* 1962; 53:295–326.
- Blume JD. Likelihood methods for measuring statistical evidence. *Stat in Med.* 2002; 21:2563–2599. [PubMed: 12205699]
- Cavalli-Sforza, LL.; Bodmer, WF. *The Genetics of Human Populations.* San Francisco: WH Freeman; 1971.
- Elston RC. Man bites dog? The validity of maximizing lod scores to determine mode of inheritance. *Am J Med Genet.* 1989; 34:487–488. [PubMed: 2624257]
- Feller, W. *An Introduction to Probability Theory and Its Applications.* 3. Vol. I. Wiley; New York: 1950.
- Greenberg DA, Abreu PC. Determining trait locus position from multipoint analysis: accuracy and power of three different statistics. *Genetic Epidemiol.* 2001; 21:299–314.
- Hacking, I. *Logic of Statistical Inference.* Cambridge; New York: 1995.
- Hodge SE, Abreu PC, Greenberg DA. Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. *Am J Hum Genet.* 1997; 60:217–227. [PubMed: 8981965]
- Hodge SE, Elston RC. Lods, wrods, and mods: the interpretation of lod scores calculated under different models. *Genet Epidemiol.* 1994; 11:329–342. [PubMed: 7813895]
- Ott, J. *Analysis of Human Genetic Linkage.* 3. Johns Hopkins; Baltimore: 1999. Multipoint linkage analysis. Chapter 6
- Risch N, Giuffra L. Model misspecification and multipoint linkage analysis. *Hum Hered.* 1992; 42:77–92. [PubMed: 1555848]
- Royall, R. *Statistical Evidence: A Likelihood Paradigm.* Chapman & Hall; London: 1994.
- Royall R. On the probability of observing misleading statistical evidence. *J Am Stat Assoc.* 2000; 95:760–780.
- Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc.* 1987; 82:605–610.

- Sham P. *Statistics in Human Genetics*. Wiley, New York, 1998; Section 3.3, “Map functions.” Smith CAB. The detection of linkage in human genetics. *J Roy Stat Soc B*. 1953; 15:153–192.
- Strauch K. MOD-score analysis with simple pedigrees: An overview of likelihood-based linkage methods. *Hum Hered*. 2007; 64:192–202. [PubMed: 17536213]
- Strug LJ, Hodge SE. An alternative foundation for the planning and evaluation of linkage analysis. I. Decoupling “error probabilities” from “measures of evidence. *Hum Hered*. 2006; 61:166–188. [PubMed: 16865000]
- Vieland VJ, Hodge SE. Review of *Statistical Evidence: A Likelihood Paradigm*, by R. Royall. *Am J Hum Genet*. 1998; 63:283–289.
- Xing C, Elston RC. Distribution and magnitude of type I error of model-based multipoint lod scores: Implications for multipoint mod scores. *Genet Epidemiol*. 2006; 30:447–458. [PubMed: 16685712]

APPENDIX

A.1. Derivation of multipoint lod score function for two marker loci

We write out the multipoint (MP) lod score function for a three-point linkage analysis involving two marker loci (A and B) and a trait locus T . We assume A and B are reasonably close together, with known distance $d \leq 0.1$ M, for example, and in linkage equilibrium. We seek to estimate the position of T , which may lie between the two marker loci or to either side. We express the MP lod function in terms of the numbers of T 's recombinants with both loci (n_{11}), with A only (n_{10}), with B only (n_{01}), and with neither locus (n_{00}). The total sample size is $N = n_{11} + n_{10} + n_{01} + n_{00}$. With similar notation, let g_{11} denote the probability of T 's recombination with both marker loci; and similarly for g_{10} , g_{01} , and g_{00} [Ott, 1999]. Let θ_{AT} , θ_{TB} , and θ_{AB} represent the recombination fractions between, respectively, A and T , T and B , and A and B . (Note that θ_{AB} is known, since d is known; see eq. A-6 below.)

Then

$$\begin{aligned} g_{11} &= (1/2)(\theta_{AT} + \theta_{TB} - \theta_{AB}), \\ g_{10} &= (1/2)(\theta_{AT} - \theta_{TB} + \theta_{AB}), \\ g_{01} &= (1/2)(-\theta_{AT} + \theta_{TB} + \theta_{AB}) \\ g_{00} &= 1 - (1/2)(\theta_{AT} + \theta_{TB} + \theta_{AB}) \end{aligned} \quad (\text{A-1})$$

[Ott, 1999]. Probabilities under the null hypothesis ($\theta_{AT} = \theta_{TB} = 1/2$) become

$$\begin{aligned} g_{11}^{(0)} &= g_{00}^{(0)} = (1/2)(1 - \theta_{AB}) \\ g_{10}^{(0)} &= g_{01}^{(0)} = (1/2)\theta_{AB}. \end{aligned} \quad (\text{A-2})$$

The relationships in (A-1) and (A-2) hold whether T is between the two marker loci or to either side.

The MP lod score, as a function of position ξ , is based on the following multinomial likelihood:

$$\begin{aligned} \text{MPLod}(\xi) &= \log_{10}[L(\xi)/L(\xi|H_0)] \\ &= n_{11} \log(g_{11}/g_{11}^{(0)}) + n_{10} \log(g_{10}/g_{10}^{(0)}) + n_{01} \log(g_{01}/g_{01}^{(0)}) + n_{00} \log(g_{00}/g_{00}^{(0)}) \end{aligned} \quad (\text{A-3})$$

where “log” always indicates common logs (base 10). Use (A-1) and (A-2) to express (A-3) in terms of the θ s:

$$\begin{aligned} \text{MPLod}(\xi) = & n_{11} \log \frac{(1/2)(\theta_{AT} + \theta_{TB} - \theta_{AB})}{(1/2)(1 - \theta_{AB})} + n_{10} \log \frac{(1/2)(\theta_{AT} - \theta_{TB} + \theta_{AB})}{(1/2)\theta_{AB}} \\ & + n_{01} \log \frac{(1/2)(-\theta_{AT} + \theta_{TB} + \theta_{AB})}{(1/2)\theta_{AB}} + n_{00} \log \frac{1 - (1/2)(\theta_{AT} + \theta_{TB} + \theta_{AB})}{(1/2)(1 - \theta_{AB})} \end{aligned} \quad (\text{A-4})$$

To formulate the MP lod as an explicit function of ξ , we need a map function that relates θ to ξ . We use the Haldane map function, which assumes independence of recombination fractions (i.e., no interference). If one considers any three linked loci, $K1$, $K2$, and $K3$, with $K2$ located between $K1$ and $K3$ on a chromosome, the Haldane assumption says that the recombinations between $K1$ and $K2$ are independent of those between $K2$ and $K3$, i.e.,

$$\text{Prob}[(\text{recomb. between } K1 \text{ and } K2) \cap (\text{recomb. between } K2 \text{ and } K3)] = \theta_{12}\theta_{23}.$$

This in turn implies

$$\theta_{13} = \theta_{12} + \theta_{23} - 2\theta_{12}\theta_{23} \quad (\text{A-5})$$

[Sham, 1998]. The corresponding map function $h(\xi)$ and its inverse $f(\theta)$ are

$$\theta = h(\xi) = (1/2)(1 - e^{-2\xi}) \quad (\text{A-6})$$

$$\xi = f(\theta) = - (1/2) \ln(1 - 2\theta), \quad (\text{A-7})$$

for ξ in Morgans.

Without loss of generality, we stipulate that locus A is located at $\xi = 0$ and B , at $\xi = d$, where d is the map distance between A and B . Thus, if T lies between the two marker loci, its position corresponds to $0 < \xi < d$ (i.e., T is in region W , as defined in Sec. 3.4, Example 2); T in region L corresponds to $\xi < 0$, and T in region R corresponds to $\xi > d$. Note that $\theta_{AB} = h(d)$ in all three regions, but the relationships between ξ and the two recombination fractions involving T depend on which region ξ is in:

$$\begin{aligned} \text{for } \xi < 0 \text{ (region } L\text{): } & \theta_{AT}^{(L)} = h(-\xi), \theta_{TB}^{(L)} = h(d - \xi) \\ \text{for } 0 < \xi < d \text{ (region } W\text{): } & \theta_{AT}^{(W)} = h(\xi), \theta_{TB}^{(W)} = h(d - \xi) \end{aligned} \quad (\text{A-8})$$

$$\text{for } \xi > d \text{ (region } R\text{): } \theta_{AT}^{(R)} = h(\xi), \theta_{TB}^{(R)} = h(\xi - d).$$

Using the information in (A-8) and substituting (A-6) into (A-4), yields (after algebraic manipulation) the multipoint lod function of ξ :

$$\begin{aligned}
\text{MPLod}(\xi) = & n_{11} \log\{[1 - (e^{2\xi} + e^{-2(d-\xi)} - e^{-2d})]/(1 + e^{-2d})\} + n_{10} \log\{[1 - (e^{2\xi} - e^{-2(d-\xi)} + e^{-2d})]/(1 - e^{-2d})\} \\
& + n_{01} \log\{[1 - (-e^{2\xi} + e^{-2(d-\xi)} + e^{-2d})]/(1 - e^{-2d})\} + n_{00} \log\{[1 + (e^{2\xi} + e^{-2(d-\xi)} + e^{-2d})]/(1 + e^{-2d})\} \text{ for } \xi < 0 \\
& n_{11} \log\{[1 - (e^{-2\xi} + e^{-2(d-\xi)} - e^{-2d})]/(1 + e^{-2d})\} + n_{10} \log\{[1 - (e^{-2\xi} - e^{-2(d-\xi)} + e^{-2d})]/(1 - e^{-2d})\} \\
& + n_{01} \log\{[1 - (-e^{-2\xi} + e^{-2(d-\xi)} + e^{-2d})]/(1 - e^{-2d})\} + n_{00} \log\{[1 + (e^{-2\xi} + e^{-2(d-\xi)} + e^{-2d})]/(1 + e^{-2d})\} \text{ for } 0 < \xi < d \\
& n_{11} \log\{[1 - (e^{-2\xi} + e^{-2(\xi-d)} - e^{-2d})]/(1 + e^{-2d})\} + n_{10} \log\{[1 - (e^{-2\xi} - e^{-2(\xi-d)} + e^{-2d})]/(1 - e^{-2d})\} \\
& + n_{01} \log\{[1 - (-e^{-2\xi} + e^{-2(\xi-d)} + e^{-2d})]/(1 - e^{-2d})\} + n_{00} \log\{[1 + (e^{-2\xi} + e^{-2(\xi-d)} + e^{-2d})]/(1 + e^{-2d})\} \text{ for } \xi > d.
\end{aligned} \tag{A-9}$$

A.2. Derivation of maxlods, $\hat{\theta}$, and $\hat{\xi}$

Each of the three regions has its own maxlod, $\hat{\theta}$, and $\hat{\xi}$, which we now derive.

A.2.1. For regions L and R

Begin with region L, where T is to the left of both markers. This means that A lies between T and B , so the independence assumption in (A-5) implies

$$\theta_{TB}^{(L)} = \theta_{AT}^{(L)} + \theta_{AB} - 2\theta_{AT}^{(L)}\theta_{AB}.$$

Substituting this expression for $\theta_{TB}^{(L)}$ into the MPLod function in (A-4) and simplifying the algebra yields the MP lod function in region L:

$$\text{MPLod}^{(L)}(\xi) = (n_{11} + n_{10}) \log \theta_{AT}^{(L)} + (n_{01} + n_{00}) \log(1 - \theta_{AT}^{(L)}) + N \log(2), \tag{A-10}$$

This is a binomial log-likelihood, with the usual constraint for linkage analysis that recombination fraction cannot exceed 0.5. If $(n_{11} + n_{10})/N < 1/2$:

$$\begin{aligned}
\widehat{\theta}_{AT}^{(L)} &= (n_{11} + n_{10})/N, \\
\text{MaxLod}^{(L)} &= (n_{11} + n_{10}) \log \widehat{\theta}_{AT}^{(L)} + (n_{01} + n_{00}) \log(1 - \widehat{\theta}_{AT}^{(L)}) + N \log(2) \\
&= (n_{11} + n_{10}) \log(n_{11} + n_{10}) + (n_{01} + n_{00}) \log(n_{01} + n_{00}) + N \log(2/N) \\
\widehat{\xi}^{(L)} &= 0 - f(\widehat{\theta}_{AT}^{(L)}) = (1/2) \ln[1 - 2(n_{11} + n_{10})/N],
\end{aligned} \tag{A-11}$$

where we used (A-7) to convert $\hat{\theta}$ to $\hat{\xi}$. If $(n_{11} + n_{10})/N \geq 1/2$:

$$\begin{aligned}
\widehat{\theta}_{AT}^{(L)} &= 0.5 \\
\text{MaxLod}^{(L)} &= 0.0. \\
\widehat{\xi}^{(L)} &\rightarrow -\infty.
\end{aligned} \tag{A-12}$$

Similar reasoning applies to region R, yielding, for $(n_{11} + n_{01})/N < 1/2$:

$$\begin{aligned}
\widehat{\theta}_{TB}^{(R)} &= (n_{11} + n_{01})/N, \\
\text{MaxLod}^{(R)} &= (n_{11} + n_{01}) \log(n_{11} + n_{01}) + (n_{10} + n_{00}) \log(n_{10} + n_{00}) + N \log(2/N) \\
\widehat{\xi}^{(R)} &= d + f(\widehat{\theta}_{TB}^{(R)}) = d - (1/2) \ln[1 - 2(n_{11} + n_{01})/N],
\end{aligned} \tag{A-13}$$

and for $(n_{11} + n_{01})/N \geq 1/2$:

$$\begin{aligned}\widehat{\theta}_{TB}^{(R)} &= 0.5 \\ \text{MaxLod}^{(R)} &= 0.0. \\ \widehat{\xi}^{(R)} &\rightarrow \infty.\end{aligned}\tag{A-14}$$

Note: The binomial form of (A-10) reveals that in region L , information about $\theta_{AT}^{(L)}$ is contained only in the number of recombinations between A and T (i.e., in the sum $n_{11} + n_{10}$) and is not affected by the number of recombinations between T and B . This is due to the independence assumption (A-5). Similarly, the sum $n_{11} + n_{01}$ contains all the information about $\theta_{TB}^{(R)}$ in region R .

A.2.2. For region W

Under the Haldane assumption, with T between A and B , $g_{11} = \theta_{AT}^{(W)} \theta_{TB}^{(W)}$, $g_{10} = \theta_{AT}^{(W)}(1 - \theta_{TB}^{(W)})$, etc. Since d is small, the product $\theta_{AT}^{(W)}\theta_{TB}^{(W)}$ is small relative to either $\theta_{AT}^{(W)}$ or $\theta_{TB}^{(W)}$; also, $\theta_{AB} \approx \theta_{AT}^{(W)} + \theta_{TB}^{(W)}$. So we can write these equations and approximations:

$$\begin{aligned}g_{11} &= \theta_{AT}^{(W)} \theta_{TB}^{(W)} \\ g_{10} &= \theta_{AT}^{(W)}(1 - \theta_{TB}^{(W)}) \approx \theta_{AT}^{(W)} \\ g_{01} &= (1 - \theta_{AT}^{(W)})\theta_{TB}^{(W)} \approx \theta_{TB}^{(W)} \\ g_{00} &= (1 - \theta_{AT}^{(W)})(1 - \theta_{TB}^{(W)}) = 1 - (\theta_{AT}^{(W)} + \theta_{TB}^{(W)}) + \theta_{AT}^{(W)}\theta_{TB}^{(W)} \approx 1 - \theta_{AB}.\end{aligned}\tag{A-15}$$

Substituting these expressions into (A-4) yields:

$$\begin{aligned}\text{MPLod}^{(W)}(\xi) &\approx (n_{11} + n_{10})\log\theta_{AT}^{(W)} + (n_{11} + n_{01})\log\theta_{TB}^{(W)} + n_{00}\log(1 - \theta_{AB}) \\ &\quad - (n_{11} + n_{00})\log[1/2(1 - \theta_{AB})] - (n_{10} + n_{01})\log(1/2\theta_{AB}) \\ &= (n_{11} + n_{10})\log\theta_{AT}^{(W)} + (n_{11} + n_{01})\log\theta_{TB}^{(W)} - n_{11}\log(1 - \theta_{AB}) - (n_{10} + n_{01})\log\theta_{AB} + N\log 2.\end{aligned}$$

This is an approximate binomial log-likelihood in the parameter $\theta_{AT}^{(W)}/\theta_{AB}$ (since θ_{AB} is assumed known, and $\theta_{TB}^{(W)} \approx \theta_{AB} - \theta_{AT}^{(W)}$). For ease of notation, let $\alpha = n_{11} + n_{10}$ and $\beta = n_{11} + n_{01}$. Then

$$\widehat{\theta}_{AT}^{(W)} \approx \frac{\alpha}{\alpha + \beta} \theta_{AB}, \quad \widehat{\theta}_{TB}^{(W)} \approx \frac{\beta}{\alpha + \beta} \theta_{AB}, \quad \text{and} \quad \widehat{\xi}^{(W)} \approx \frac{\alpha}{\alpha + \beta} d,\tag{A-16}$$

and the corresponding maxlod becomes:

$$\text{MaxLod}^{(W)}(\xi) \approx \alpha \log[\alpha/(\alpha + \beta)] + \beta \log[\beta/(\alpha + \beta)] + n_{11} \log[\theta_{AB}^2/(1 - \theta_{AB})] + N \log 2.\tag{A-17}$$

We have confirmed numerically that for $d \leq 0.1$, these approximations in (A-16) and (A-17) are very close to the exact values.

Special case: If $n_{11} + n_{10} + n_{01} = 0$, the MP lod curve is bimodal, as in Fig. 2f. The two maxima occur at $\xi = 0$ and $\xi = d$, both with the max MPLod value of $N \log 2$.

A.2.3. Possible “ties”

Ties can occur in one of three ways: If $n_{11} + n_{10} = 0$ and $n_{01} > 0$, there are recombinations with B but none with A (as in Fig. 2e). In this case, $\text{MPLod}(\xi)$ is defined for $\xi = 0$, and we count both $\text{MaxLod}^{(W)}$ and $\text{MaxLod}^{(L)}$ as tied for global maximum, with the common value $N \log 2$. Similarly, if $n_{11} + n_{01} = 0$ and $n_{10} > 0$, $\text{MaxLod}^{(W)}$ and $\text{MaxLod}^{(R)}$ are tied, with the same common value of $N \log 2$. Finally, if $n_{11} + n_{10} + n_{01} = 0$ (Fig. 2f), we count all three maxima as tied, with, again, the common value of $N \log 2$.

A.3. Calculating probabilities of type I error

There are $(N+3)!/(N!3!)$ distinct ways (configurations) to distribute N items among four distinct categories [Feller, 1950]. For each given sample size N , we determined all possible configurations $(n_{11}, n_{10}, n_{01}, n_{00})$, and we calculated the probability of observing each configuration under the null hypothesis that the trait is not linked to either A or B . We also calculated, for each configuration, the three maxima outlined above: $\text{MaxLod}^{(L)}$ in (A-11) or (A-12), $\text{MaxLod}^{(R)}$ in (A-13) or (A-14), and $\text{MaxLod}^{(W)}$ in (A-17). If none of these three maxima reached our designated cutoff value (e.g., 0.9 or 1.5), we discarded that configuration. If at least one did reach the cutoff value, we determined which of the three was the *global* maximum (or which two, or all three, in the case of ties). We tabulated the probabilities of all seven possible outcomes – as described in Sec. 3.4, Example 2 – and these results are shown in Table 5.

Examples. Let $N = 10$, $d = 0.1$, and consider a cutoff value of 0.9. The configuration $(n_{11}, n_{10}, n_{01}, n_{00}) = (3, 2, 3, 2)$ in Fig. 2a occurs with probability 9.360×10^{-5} . For this configuration, none of the three maxima exceeds the cutoff, so the configuration is discarded. The configuration $(0, 1, 5, 4)$ in Fig. 2b occurs with probability 4.664×10^{-7} . Both $\text{MaxLod}^{(L)}$ and $\text{MaxLod}^{(W)}$ exceed 0.9, and of these two, $\text{MaxLod}^{(W)}$ is larger, so the probability of 4.664×10^{-7} gets added to the “W” column in Table 5. Configuration $(0, 0, 0, 10)$ in Fig. 2f occurs with probability 3.776×10^{-4} and gets added to the “LWR” column.

Note: The MLEs in (A-16) and $\text{MaxLod}^{(W)}$ in (A-17) are based on the approximations in (A-15) but are very close to their exact values, since d is small. Hence we refer to these as “exact calculations,” in order to distinguish them from the simulation results in this study.

A.4. Behavior of MP lod score function for two loci very close together

Consider the situation where $d \rightarrow 0$. This implies $\theta_{AT} \approx \theta_{TB}$, which we will denote by the common value θ ; also $\theta_{AB} \rightarrow 0$. Therefore, in (A-1), $g_{11} \rightarrow \theta$, $g_{00} \rightarrow 1 - \theta$, and g_{10} and $g_{01} \rightarrow 0$. In (A-2), $g_{11}^{(0)}$ and $g_{00}^{(0)} \rightarrow 1/2$, whereas $g_{10}^{(0)}$ and $g_{01}^{(0)} \rightarrow 0$. We assume that $n_{10} = n_{01} = 0$, so $N = n_{11} + n_{00}$. Inserting these limits into the formulas in (A-3) yields

$$\begin{aligned} \text{MPLod}(\theta) \rightarrow & \\ n_{11} \log [\theta/(1/2)] + n_{00} \log [(1 - \theta)/(1/2)] & \text{ for } \theta \in (0, 1/2] \\ -\infty \text{ if } n_{11} > 0, \text{ or } N \log 2 \text{ if } n_{11} = 0 & \text{ for } \theta = 0 \end{aligned}$$

which corresponds exactly to the two-point lod score function.

Figure 1. Observed type I error rates for cutoff value of 0.9, evaluated at a single position, in 1,000 datasets, as function of dataset size. Data were generated assuming either dominant (left graphs) or recessive (right graphs) inheritance, with reduced penetrance as indicated. Top two graphs show lods, bottom two show mods.

Figure 2.

Examples of multipoint lod score curves and the datasets that can yield them, with $d = 0.1$ M. Horizontal axis gives position ξ , in Morgans; vertical axis gives MP lod score as a function of ξ . Locus A is located at $\xi = 0$, and B is at $\xi = 0.1$. The two vertical lines (vertical axis and line at $\xi = 0.1$) are asymptotes, except when there is no recombination with the corresponding locus. For each part of the figure, the four numbers in parentheses indicate the configuration $(n_{11}, n_{10}, n_{01}, n_{00})$, where n_{11} = number of meioses resulting from recombination with both loci, n_{10} = number resulting from recombination with A only, etc. See Example 2, Sec. 3.4, for more detailed explanation.

Figure 3. Observed type I error rates for cutoff value of 0.9, maximized over 10 cM span (11 marker loci), in 1,000 datasets, as function of dataset size. Data were generated assuming either dominant (left graphs) or recessive (right graphs) inheritance, with reduced penetrance as indicated. Top two graphs show lods, bottom two show mods.

Figure 4.

Observed type I errors of MP lods maximized over different spans, as indicated, in 1,000 datasets, as function of dataset size, for cutoff value of 0.9. Data were generated assuming either dominant (left graph) or recessive (right graph) inheritance, with 50% penetrance.

Table 1

Observed Type I Errors for Lods and Mods Evaluated at a Single Position* in 1,000 Datasets, as Function of Dataset Size, for Cutoff Values of 0.9 and 1.5.

(a) Lods			
No. of fams/dataset			
GM	10	20	30
D20	.022	.010	.001
D50	.013	.002	.000
D80	.002	.000	.000
R20	.013	.000	.000
R50	.004	.000	.000
R80	.001	.000	.000

(b) Mods						
No. of fams/dataset						
GM	10	20	30	40	50	75
D20	.047/.010	.031/.011	.018/.004	.010/.005	.014/.005	.004/.001
D50	.045/.011	.015/.005	.007/.006	.004/.001	.001/.001	.000/.000
D80	.028/.008	.004/.001	.005/.002	.000/.000	.000/.000	.000/.000
R20	.034/.009	.034/.008	.022/.006	.020/.001	.018/.006	.007/.002
R50	.040/.015	.027/.005	.024/.005	.011/.004	.003/.001	.011/.007
R80	.037/.008	.025/.005	.027/.008	.018/.006	.016/.006	.006/.003

In part (a), results are shown only for cutoff value of 0.9. In part (b), first number (plain font) in each cell gives result for the 0.9 cutoff, second number (italics) gives results for 1.5.

* This position was located halfway between two markers separated by 10 cM.

Table 2

Observed Type I Errors for Lods and Mods in 1,000 Datasets, Maximized over 10 cM,* as Function of Dataset Size, for Cutoff Values of 0.9 and 1.5.

(a) Lods					
No. of fams/dataset					
GM	10	20	30	40	
D20	.048/.013	.018/.008	.001/.000	.001/.001	
D50	.024/.013	.001/.001	.001/.001	.000/.000	
D80	.005/.002	.000/.000	.000/.000	.000/.000	
R20	.018/.003	.001/.000	.000/.000	.000/.000	
R50	.007/.005	.001/.000	.000/.000	.000/.000	
R80	.002/.001	.000/.000	.000/.000	.000/.000	

(b) Mods						
No. of fams/dataset						
GM	10	20	30	40	50	75
D20	.098/.027	.054/.021	.036/.009	.014/.003	.022/.007	.012/.005
D50	.069/.031	.021/.009	.013/.003	.002/.002	.003/.002	.000/.000
D80	.052/.023	.015/.008	.005/.004	.000/.000	.000/.000	.000/.000
R20	.081/.025	.071/.027	.066/.031	.044/.014	.024/.007	.018/.007
R50	.083/.024	.069/.019	.066/.024	.040/.012	.029/.006	.018/.008
R80	.068/.025	.072/.030	.054/.024	.039/.013	.033/.010	.015/.004

The first number (plain font) in each cell gives results for the 0.9 cutoff, second number (italics) gives results for 1.5.

*The 10 cM encompassed 11 equispaced markers; lods and mods were evaluated at 51 equispaced points.

Table 3

Observed Type I Errors for Lods Maximized over a Larger Span, in 1,000 Datasets, as Function of Dataset Size, for Cutoff Values of 0.9 and 1.5.

		No. of fams/dataset				
	GM	10	20	30	40	50
Maximization over						
18 markers, 50 cM	D50	.048/.026	.006/.002	.003/.003	.001/.001	.000/.000
O	R50	.014/.005	.001/.001	.000/.000	.000/.000	.000/.000
	D50	.087/.041	.016/.009	.003/.003	.000/.000	.000/.000
O	R50	.027/.016	.000/.000	.002/.001	.000/.000	.000/.000

The first number (plain font) in each cell gives results for the 0.9 cutoff, second number (italics) gives results for 1.5.

Table 4

Observed Type I Errors for Two-Point Lods Maximized Over θ 0 [0.0, 0.2] in 2,000 Datasets, as Function of Dataset Size.

Cutoff value ^I	No. of fams/dataset				
	10	20	30	40	50
0.59 (.05)	.0575	.0330	.0200	.0095	.0070
0.90 (.02)	.0275	.0190	.0130	.0060	.0050
1.17 (.01)	.0135	.0110	.0070	.0035	.0030

^I Number in parentheses gives the approximate nominal type I error probability (one-sided) associated with that cutoff value.

Table 5

Probabilities* of MaxMPLod ≥ 0.9 Occurring in the Specified Region, as Function of Sample Size *N*, Under Null Hypothesis of No Linkage with Either Locus

(a) <i>d</i> = 10 cM							
<i>N</i>	<i>L</i>	<i>W</i>	<i>R</i>	<i>LW</i>	<i>LR</i>	<i>WR</i>	<i>LWR</i>
5	.0000	.0039	.0000	.0118	.0000	.0118	.0194
10	.0051	.0007	.0051	.0006	.0038	.0006	.0004
15	.0104	.0001	.0104	.0000	.0057	.0000	.0000
20	.0129	.0000	.0129	.0000	.0056	.0000	.0000
25	.0140	.0000	.0140	.0000	.0050	.0000	.0000
(b) <i>d</i> = 5 cM							
<i>N</i>	<i>L</i>	<i>W</i>	<i>R</i>	<i>LW</i>	<i>LR</i>	<i>WR</i>	<i>LWR</i>
5	.0000	.0012	.0000	.0068	.0000	.0068	.0245
10	.0033	.0002	.0033	.0004	.0060	.0004	.0006
15	.0071	.0000	.0071	.0000	.0092	.0000	.0000
20	.0093	.0000	.0093	.0000	.0092	.0000	.0000
(c) <i>d</i> = 1 cM							
<i>N</i>	<i>L</i>	<i>W</i>	<i>R</i>	<i>LW</i>	<i>LR</i>	<i>WR</i>	<i>LWR</i>
5	.0000	.0001	.0000	.0015	.0000	.0015	.0297
10	.0008	.0000	.0008	.0001	.0088	.0001	.0009
15	.0020	.0000	.0020	.0000	.0152	.0000	.0000
20	.0028	.0000	.0028	.0000	.0171	.0000	.0000

* Exact calculations, not simulations

L = Left, *W* = Between, *R* = Right; probabilities in **bold** contribute to “observed” type I error rate; see text for fuller explanation.

N = number of informative meioses

d = distance between the two marker loci

Table 6

Total (“Theoretical”) Type I Error Levels* as a Function of N and d for Cutoff values 0.9 and 1.5

(a) Cutoff value = 0.9					
N	d (in cM)				
	10	5	1	0.1	0.01
10	.016	.014	.012	.011	.011
20	.031	.028	.023	.021	.021
30	.033	.029	.024	.022	.021
40	.030	.027	.022	.020	.019
50	.026	.023	.019	.017	.016
60	.040	.036	.030	.026	.026
70	.025	.029	.024	.021	.021
80	.026	.023	.019	.017	.017
90	.034	.031	.026	.023	.022
100	.027	.025	.021	.018	.018
(b) Cutoff value = 1.5					
N	d (in cM)				
	10	5	1	0.1	0.01
10	.016	.014	.012	.011	.011
20	.009	.008	.007	.006	.006
30	.004	.004	.003	.003	.003
40	.005	.005	.004	.003	.003
50	.005	.005	.004	.003	.003
60	.005	.005	.004	.003	.003
70	.009	.008	.007	.006	.006
80	.008	.007	.006	.005	.005
90	.007	.006	.005	.004	.004
100	.005	.005	.004	.003	.003

* Exact calculations, not simulations
 N = number of informative meioses

p = distance between the two marker loci

Table 7

Distribution of MPMOD Scores Attaining Cutoff values 0.9 and 1.5, Evaluated at a Single Position*, out of 1,000 Datasets.

(a) 20 fams/DS													
AM	GM = D20		GM = D50		GM = D80		GM = R20		GM = R50		GM = R80		
	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	
D 1-3	8	3	3	2	1	0	11	1	11	1	12	3	
D 4-6	1	1	1	0	0	0	20	6	14	3	11	2	
D 7-9	0	0	0	0	0	0	0	0	1	1	0	0	
R 1-3	22	7	11	3	3	1	3	1	1	0	2	0	
R 4-6	0	0	0	0	0	0	0	0	0	0	0	0	
R 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
totals	31	11	15	5	4	1	34	8	27	5	25	5	
(b) 40 fams/DS													
AM	GM = D20		GM = D50		GM = D80		GM = R20		GM = R50		GM = R80		
	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	
D 1-3	0	0	1	0	0	0	17	0	9	3	14	4	
D 4-6	0	0	0	0	0	0	3	1	2	1	4	2	
D 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
R 1-3	10	5	3	1	0	0	0	0	0	0	0	0	
R 4-6	0	0	0	0	0	0	0	0	0	0	0	0	
R 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
totals	10	5	4	1	0	0	20	1	11	4	18	6	

* This position was located halfway between two markers separated by 10 cM.

See text for fuller explanation of this table.

Table 8

Distribution of MaxMPMOD Scores Attaining Cutoff values 0.9 and 1.5, Maximized over 10 cM,* out of 1,000 Datasets.

(a) 20 fams/DS													
AM	GM = D20		GM = D50		GM = D80		GM = R20		GM = R50		GM = R80		
	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	
D 1-3	20	11	5	3	5	3	40	8	47	11	44	10	
D 4-6	1	1	0	0	0	0	30	19	20	6	26	18	
D 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
R 1-3	33	9	16	6	10	5	1	0	2	2	2	2	
R 4-6	0	0	0	0	0	0	0	0	0	0	0	0	
R 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
totals	54	21	21	9	15	8	71	27	69	19	72	30	
(b) 40 fams/DS													
AM	GM = D20		GM = D50		GM = D80		GM = R20		GM = R50		GM = R80		
	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	≥0.9	≥1.5	
D 1-3	1	1	0	0	0	0	42	12	35	9	33	8	
D 4-6	0	0	0	0	0	0	2	2	5	3	6	5	
D 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
R 1-3	13	2	2	2	0	0	0	0	0	0	0	0	
R 4-6	0	0	0	0	0	0	0	0	0	0	0	0	
R 7-9	0	0	0	0	0	0	0	0	0	0	0	0	
totals	14	3	2	2	0	0	44	14	40	12	39	13	

* The 10 cM encompassed 11 equispaced markers; mods were evaluated at 51 equispaced points.
See text for fuller explanation of this table.