Am. J. Hum. Genet. 50:598–606, 1992

Comparison of a Multipoint Identity-by-Descent Method with Parametric Multipoint Linkage Analysis for Mapping Quantitative Traits

David E. Goldgar and Rebecca S. Oniki

Department of Medical Informatics, University of Utah, Salt Lake City

Summary

We previously developed a method of partitioning genetic variance of a quantitative trait to loci in specific chromosomal regions. In this paper, we compare this method—multipoint IBD (identical by descent) method (MIM)—with parametric multipoint linkage analysis (MLINK). A simulation study was performed comparing the methods for the major-locus, mixed, and two-locus models. The criterion for comparisons between MIM and MLINK was the average lod score from multiple replicates of simulated data sets. The effect of gene frequency, dominance, model misspecification, marker spacing, and informativeness are also considered in a smaller set of simulations. Within the context of the models examined, the MIM approach was found to be comparable in power with parametric multipoint linkage analysis when (*a*) parental data are unknown, (*b*) the effect of the major locus is small and there is additional genetic variation, or (*c*) the parameters of the major-locus model are misspecified. The performance of the MIM method relative to MLINK was markedly lower when the allele frequency at the trait locus was .2 versus .5, particularly for the case when parental data were assumed to be known. Dominance at the trait major locus, as well as marker spacing and heterozygosity, did not appear to have a large effect on the ELOD comparisons.

Introduction

As a result of the GENOME initiative, a set of highly polymorphic index markers located 10–15 cM apart are being developed for each human chromosome. Such an index map may now make it possible to localize genes contributing to multifactorial quantitative traits or complex diseases which do not exhibit a simple Mendelian pattern of inheritance. In humans most previous efforts at linkage of complex diseases and traits have utilized sib-pair (Penrose 1935; Suarez et al. 1978; Lange 1986) or affected-relative-pair (Weeks and Lange 1988) methods. For quantitative traits, Haseman and Elston (1972) proposed a sib-pair method based on the regression of the squared sib-pair difference for the trait on their estimated genetic correlation at the marker locus. This method was extended (Amos and Elston 1989; Amos et al. 1989) and was used to examine linkage of a number of hypertension-related quantitative traits to a number of blood-group and electrophoretic markers (Wilson et al. 1991). Our approach also uses the concept of identical by descent (IBD) but is based on the amount of genetic material shared in a given chromosomal region rather than at a single marker locus. Specifically, we estimate the expected proportion of genetic material in a particular chromosomal region shared IBD for each pair of siblings in a given sibship, on the basis of their genotypes at a series of marker loci in that region. These estimates are then used to partition genetic variance of a human quantitative trait to loci located in that region. This method will be denoted the multipoint IBD method (MIM) throughout the remainder of the present paper.

If the quantitative trait under investigation displays an inheritance pattern indicative of a major gene segre-

Received July 25, 1991; final revision received November 5, 1991.

Address for correspondence and reprints: David E. Goldgar, Ph.D., Genetic Epidemiology, 420 Chipeta Way, #180, Salt Lake City, UT 84108.

0002-9297/92/5003-0020$02.00

gating in the families, and if this hypothesized major locus can be accurately characterized in terms of its model parameters, then the parametric methods of pairwise linkage analysis (Morton 1955; Ott 1974) and multipoint linkage analysis (Lathrop et al. 1984, 1985) may be used. In addition to being the most powerful test for linkage if the model is correctly specified, the parametric approach has the advantage of providing precise estimates of the hypothesized major gene's location along the genetic map. Although a number of investigators have examined the power of the lod-score method to detect linkage between a quantitative-trait locus and a single marker (Lange et al. 1976; Demenais et al. 1988; Boehnke 1990), no studies have been performed examining the power of this method in the multipoint case. It is important to point out that the lod-score approach is only applicable when there is significant evidence of a major locus influencing the trait; moreover, a genetic model of this locus must be specified. However, if the model is misspecified, this may lead to lower power, false positives, and biases in the estimation of the recombination fraction(s). Although some studies have shown that accurate characterization of the genetic model is not necessarily a requirement for detecting or excluding linkage (Risch et al. 1989; Skolnick et al. 1989), it is not known how general this effect is. For example, Clerget-Darpoux et al. (1986) found that misspecification of the degree of dominance could cause a reduction in the expected lod score, particularly for a recessive disease with moderately high (.50) penetrance.

Our initial studies (Goldgar 1990; Goldgar and Oniki 1990) have shown the MIM method to be considerably more powerful than the Haseman-Elston method, even when the latter was modified to examine multiple loci. Of perhaps more interest is the comparison between our approach and multipoint linkage analysis. Given the relatively model-free nature of our method, if it compares favorably in terms of power with standard multipoint linkage analysis under an assumed known model of single-locus inheritance, we can be assured that under more realistic conditions, where there may be a more complex underlying etiology, the proposed method would be superior. Another potential advantage of the MIM method is that it is based only on sibship trait values; parental phenotype data are not used. This may be important for linkage studies of traits or diseases in which accurate or meaningful measurements can be obtained only in children or adolescents. One would anticipate a substantial loss of power for detecting linkage of a quantitative trait when using multipoint linkage analysis in the absence of parental phenotypes. In the present study we compare the power of our method (MIM) with that of traditional multipoint linkage analysis, for a variety of major-locus models.

Methods

MIM

A more complete description of the MIM method can be found in Goldgar's (1990) paper. In brief, the MIM method assumes the absence of genetic interference across a known map of genetic markers. We assume that the quantitative trait under study is due to additive genetic effects and a normally distributed random environmental component. The method is parameterized by two parameters: h^2, the proportion of total trait variance due to genetic influences, and P, the proportion of *genetic* variance due to loci in the test chromosomal region defined by the marker loci being studied. For each sib pair we first estimate the proportion of the test region shared by the pair IBD conditional on the marker genotypes in the region. These estimates are used to form the predicted covariance matrix of the sibship trait values as a function of these proportion IBD estimates and the parameters P and h^2. The likelihood of the standardized sibship trait values is then given by a multivariate normal density function with a mean of zero and with covariance matrix as determined above. The overall likelihood of the data set is the product of these density functions across families. Numerical maximum-likelihood techniques are used to estimate P and to test the null hypothesis $P = 0$.

MLINK

As implemented in the present study, parametric multipoint linkage analysis was performed using the MLINK program of the package LINKAGE (Lathrop et al. 1985), under the assumption of no interference across a known genetic map. The usual single-locus model—i.e., that the trait phenotype is composed of the effects of a single major locus plus normally distributed individual-specific residual variation around each major genotype mean—was assumed. The major-locus model is parameterized by five parameters: q, the allele frequency at the major locus; μ_i, the mean of the ith genotype, $i = 1, 2, 3$; and σ^2, the residual variance about each genotype mean. In addition, we have a single parameter, θ, related to the position of

the hypothesized trait locus on the fixed genetic map. For convenience this is taken to be the recombination fraction between the major locus and a flanking marker locus. The recombination fraction between the trait and other markers is determined from θ and the fixed genetic map. For fixed values of the major-locus parameters the likelihood of pedigree as a function of θ is calculated. The value of θ (location of major locus) which maximizes the likelihood can be determined and compared with the likelihood calculated under the hypothesis that the major locus is unlinked to the marker region.

The comparison of the two methods will be done through use of a simulation program which was written to evaluate the MIM method and modified and extended to implement the studies reported here. This program is described in some detail in Goldgar's (1990) paper but can be summarized as follows: The program allows for a simulation of a quantitative trait with the following components: trait loci which are located in the marker region being tested, an unlinked trait locus of relatively large effect, a polygenic component, and random environmental effects. On the basis of input parameters of allele frequency, degree of dominance, trait heritability (in the broad sense), and proportion of genetic variance due to the aggregate effects of the "major" trait locus or loci, the magnitudes of the effects of alleles at each trait locus located on the test chromosomal region are determined. To simulate the polygenic component, a similar procedure is used to determine allelic effects of 10 additive trait loci each with two equally frequent alleles, which are unlinked to the test region and to each other. The random environmental component is simulated by generating for each individual an independent random normal deviate with 0 mean and variance $1 - h^2$. The program uses an input genetic map of marker and trait loci to simulate the process of recombination between linked loci by assuming Haldane's (1919) function for translating map distances into recombination fractions. Note that this assumes the absence of interference between recombination in adjacent intervals.

Ideally, we would have liked to estimate all the relevant parameters through maximum-likelihood methods and to use as the criterion for comparison the empirical power derived from evaluation of the generalized likelihood-ratio test. However, given the computational constraints in the MLINK procedure, it was not possible either (1) to perform maximum-likelihood estimation of the location of the trait locus's with regard to the marker region or (2) to estimate the parameters of the genetic model for each simulated data set. Instead, the multipoint lod score was calculated assuming the correct position of the trait major locus. For the MIM analysis an analogous procedure was performed where we calculated the $\log_{10}$ of the ratio of the likelihood under the hypothesis of P equal to its true value to the corresponding likelihood under the null hypothesis $P = 0$. The average lod score from 50 data sets generated under the same conditions was used as the criterion for comparing the MIM and MLINK methods.

Models to Be Tested

For the majority of the comparisons in the present paper we restrict ourselves to an additive major locus with two equally frequent alleles. This allows all three genotypic distributions to be represented in reasonably high frequency in each simulated sample. The primary factors to be investigated in the simulation comparisons are the magnitude of the effect of the major locus and the nature of the residual variation. For the former, we assume that the major trait locus accounts for 30%–70% of the total phenotypic variation. Three models of the residual variance are considered: (1) all residual variation is due to individual specific random environmental effects; (2) 50% of the residual is polygenic and 50% random environment; and (3) 50% of the residual is due to a second major locus and 50% is random. The experimental design is detailed in table 1. In terms of the MIM method, the parameter P, the proportion of genetic variance due to the test marker region, is calculated as $V_{LML}/(1 - V_{RE})$ and trait heritability is equal to $(1 - V_{RE})$.

For the main body of simulations we assumed that the trait locus was located equidistant between two fully informative flanking markers located 50 cM apart. This implies a recombination fraction of .197 between the trait locus and each marker. We also assumed that the characteristics of the major gene (gene frequency, genotypic means, and common SD) that were needed for the parametric analyses were known and were set equal to their input values. Similarly, we assumed that the trait heritability needed for the MIM analysis was known as well. For each of the models examined we reanalyzed the generated data sets by the MLINK program and assuming that the parental trait data were unknown. In all the simulation reported here a data-set size of 100 families with four offspring each was assumed. Our preliminary studies show that this sample size has reasonable power for detecting the

Table 1

Major-Locus Models Examined in Primary Simulations

	Major-Locus Parameter[b]				Residual Variation[c]				
					Model I	Model II		Model III	
V_{LML}[a]	μ_1	μ_2	μ_3	σ^2	V_{RE}	V_{PG}	V_{RE}	V_{UML}	V_{RE}
.7	1.183	.0	−1.183	.3	.3	.15	.15	.15	.15
.6	1.095	.0	−1.095	.4	.4	.20	.20	.20	.20
.5	1.000	.0	−1.000	.5	.5	.25	.25	.25	.25
.4	.894	.0	−.894	.6	.6	.30	.30	.30	.30
.3	.775	.0	−.775	.7	.7	.35	.35	.35	.35

NOTE.—The major locus was assumed to have two equally frequent additive alleles.

[a] Variance due to linked major locus.

[b] μ_i = Mean of major-locus genotype i; and σ^2 = variance of each major-locus genotype.

[c] V_{RE} = variance due to random environmental effects; V_{PG} = variance due to polygenic effects; and V_{UML} = variance due to unlinked major locus.

effects of loci influencing a quantitative trait and is achievable in practice as well.

In a subset of models (those with V_{LML} = .4) we examined the effect of model misspecification by estimating the model parameters for each method, on the basis of an independent data set of 200 nuclear families. The major-locus parameters for the parametric multipoint method were estimated using the ILINK program assuming no linkage, and the overall trait-heritability estimate needed for the MIM method was estimated as twice the sibship intraclass correlation. These parameter estimates were then used for an analysis of 50 replications, as described above. Although it would have been desirable to estimate parameters for each generated data set and to use these estimates in the analysis of that particular data set, computational considerations did not permit that approach. For comparative purposes, the same data sets were analyzed using the true simulated values of the parameters.

In addition to the primary simulations and those designed to investigate model misspecification, three additional factors were investigated: (1) the effect of distance between markers (50, 25, and 10 cM) and marker heterozygosity (H = 1.0, .7, .5, and .3), (2) allele frequency, and (3) dominance (q = .2; additive, dominant, and recessive) at the major locus. For these latter experiments we chose the mixed model (model II) with V_{LML} = .4 as the standard of comparison; it seemed most similar to the kind of traits one might see in practice, and for the mixed model and two-locus model it showed nearly equivalent average lod scores for the MLINK and MIM methods. All analyses were performed on a DECstation 5000 RISC-based workstation.

Results

Table 2 displays the results of comparisons between MIM and MLINK for the main body of simulations, giving for each model and major-locus effect the average lod score calculated using MLINK and MIM and the ratio between the MIM average lod score and its equivalent obtained from MLINK. When all the residual variance was due to random normal environmental effects, the assumptions inherent in the parametric analysis are satisfied, and in all cases the MLINK analysis had higher lod scores than did the MIM analysis. When the major gene accounted for 50% or less of the total variation, the MIM method was about 75% as efficient as the MLINK approach, when average lod scores were used as the criteria. The preceding assumed that the model parameters were known without error and that the parental trait values were present. When we eliminated the parental phenotype data, the relative efficiency of the MIM approach was approximately 90% for all major-gene effects examined. However, when the model included other genetic variation in the residual (models II and III), the MIM method fared somewhat better. When parental trait data were included, the MIM analysis was roughly equivalent to the MLINK analysis, for .4 and .3 values of the linked major-locus effect, while, without parental phenotypes, MIM performed as well as or better than MLINK, for all magnitudes of the linked major-

Table 2

Power Comparisons between MIM and MLINK, as a Function of Major-Locus Effect and Model of Residual Variation

A. Parental Trait Phenotypes Known

	Model I (major locus)			Model II (mixed model)			Model III (two locus)		
	ELOD[a]			ELOD			ELOD		
V_{LML}	MLINK	MIM	Ratio[b]	MLINK	MIM	Ratio	MLINK	MIM	Ratio
.7	5.45	2.97	.55	5.96	3.88	.65	5.19	3.44	.66
.6	2.99	1.84	.61	3.25	2.41	.74	3.32	2.42	.73
.5	1.63	1.22	.75	2.14	1.72	.80	2.11	1.94	.92
.4	.88	.66	.75	1.00	.97	.97	.91	.88	.97
.3	.32	.24	.75	.52	.58	1.10	.56	.60	1.06

B. Parental Trait Phenotypes Unknown

	Model I (major locus)		Model II (mixed model)		Model III (two locus)	
V_{LML}	ELOD[c]	Ratio	ELOD	Ratio	ELOD	Ratio
.7	3.35	.89	3.97	.98	3.58	.96
.6	1.94	.94	2.27	1.06	2.41	1.00
.5	1.29	.95	1.60	1.07	1.71	1.13
.4	.72	.91	.88	1.10	.87	1.01
.3	.26	.92	.45	1.28	.51	1.18

[a] Average lod score from 50 replications.
[b] Ratio of average lod score from MIM to that from MLINK.
[c] Average lod score for MLINK. ELODs for MIM are identical to those in section A.

locus examined. In the analysis of 500 replicates for type I error when MIM was used, there were no significant deviations from the nominal level of .05, for any of the models examined.

The effect of model misspecification can be seen in table 3. Table 3*A* shows the parameter estimates obtained for the three methods. As one would expect, the estimates of the parameters of the major locus were poorer when additional genetic variation was present in the model. In particular, a greater effect for the major locus was inferred than was actually present. The estimates of heritability used in the MIM analysis were quite good for all three models. As shown in table 3*B*, when heritability was estimated instead of assumed to be equal to its true value, there was virtually no effect on the average lod score for the MIM method. However, when the estimated model parameters were used, the average lod scores when MLINK was used were reduced, most noticeably for model III (the two-locus model), in which there was a 22% reduction in average lod score.

Table 4 examines the effects, on both methods, of both a lower gene frequency and dominance at the major locus. The effect that reduced gene frequency had on the ELOD ratio was more severe than the effect of dominance at the major locus. For known parental phenotypes, there was a marked drop in the performance of the MIM method compared against the MLINK approach, for all three models, from an ELOD ratio of .97 for this model when $q = .5$ to .76 for the additive case when $q = .2$. The addition of a dominance component reduced the relative ELOD comparison only slightly from that seen in the additive case. When parental data were assumed to be unavailable, the drop-off in performance under the dominant model was not as severe (1.1 vs. 1.0), and, relative to MLINK, the MIM method actually performs better than it does in the additive, $q = .5$ case.

Table 5 shows the effects that (*a*) the distance between markers and (*b*) the informativeness of the two markers have on the two methods. Neither reduced marker informativeness nor marker distance seemed to have a consistent effect on the ELOD ratio. The lowest ratio observed (.8) occurred for the 25-cM map

Table 3

Effect of Model Misspecification on MIM-MLINK Comparisons

A. Estimated Parameter Values				
Parameter	True	Model I (major locus; h^2 = .4), Estimated	Model II (mixed; h^2 = .7), Estimated	Model III (two locus; h^2 = .7), Estimated
q	.50	.49	.56	.44
μ_1	.894	.718	.951	1.143
μ_2	.0	.124	−.084	.153
μ_3	−.894	−.924	−1.132	−.896
σ^2	.6	.675	.417	.434
h_2		.40	.69	.67

B. Comparison of ELODs by Using Estimated Parameters						
	Model I (major locus)		Model II (mixed)		Model III (two locus)	
	True	Estimated	True	Estimated	True	Estimated
ELOD:						
MIM	.64	.64	1.01	1.02	.99	1.00
MLINK	.86	.75	1.12	1.02	.98	.76
Ratio	.75	.86	.91	1.00	1.01	1.32

NOTE.—The effect of the linked locus was equal to 40% of the total phenotypic variance. Parental phenotype data were assumed to be known.

and the lowest H value, while some of the higher ratios were observed with similarly low heterozygosity but for both the 50- and 10-cM distances. Last, we note that the MIM analysis required about a tenth as much CPU time as did the MLINK likelihood calculations.

Discussion

From table 2 it is clear that the relative advantage of MIM compared with the parametric multipoint analysis is largest in those situations in which there is a large degree of uncertainty regarding the major-locus genotype, given the trait phenotype. This is true regardless of whether this uncertainty is a result of a high degree of overlap between genotype means or is due to the presence of other genetic effects not included in the parametric model. In particular, there is a large reduction in ELOD for the parametric analysis when we assumed that parental quantitative trait values are unknown. This effect was largest for model I, with the

Table 4

Effects of Gene Frequency and Degree of Dominance on Two Methods for Mixed Model (Model II) with a Linked Major Locus Accounting for 40% of Phenotypic Variance

	q = .5: Additive		q = .2					
			Additive		Dominant		Recessive	
	PPK	PPU	PPK	PPU	PPK	PPU	PPK	PPU
ELOD:								
MIM	.97	. . .	.96	. . .	1.17	. . .	.78	. . .
MLINK ...	1.00	.88	1.26	.94	1.63	1.04	1.11	.78
Ratio	.97	1.10	.76	1.02	.72	1.13	.70	1.00

NOTE.—Model parameters were fixed at their simulated values. Results are based on 100 replicates. PPK = parental phenotypes known; and PPU = parental phenotypes unknown.

Table 5

Effects That Marker Heterozygosity and Map Distance between Markers Have on Average Lod Score of Two Methods

D_a (in cM) AND H	ELOD		
	MIM	MLINK	RATIO
50:			
.3	.43	.42	1.02
.5	.52	.57	.92
.7	.64	.75	.85
1.0	.98	1.13	.87
25:			
.3	.65	.82	.80
.5	.98	1.22	.80
.7	1.27	1.46	.86
1.0	2.42	2.61	.93
10:			
.3	1.57	1.62	.97
.5	1.94	2.13	.91
.7	3.01	3.08	.98
1.0	3.43	3.52	.97

NOTE.—The simulated model was a mixed model (model II) with V_{LML}, and parental phenotype data were assumed to be known.
[a] Distance between markers.

major gene accounting for at least 50% of the total variance; in these cases a 30%–40% reduction in ELOD was observed.

It is difficult to directly compare the average lod scores obtained in our simulations with those of other investigators who have examined the power of the lod score method to detect linkage of quantitative traits. The other studies concentrated on dominant major loci, were restricted to pairwise analyses, used different family structures, and, in general, examined tighter linkage. In order to compare our results we simulated one of the models used by Boehnke—specifically, a dominant major locus with a frequency of .133, a separation between high and low genotypic means of 1.5 SDs, and complete linkage with a fully informative marker locus. Boehnke estimated for this case that a sample size of 972 individuals would be needed to achieve a lod score of 3.0, assuming randomly sampled nuclear families of size 5 and that linkage at .10 would result in an increase by a factor of 2.5, yielding an estimated sample size of 2,430 individuals for this case. For this model, we arrived at a figure of 2,200 individuals for nuclear families of size 6 and for a trait locus flanked by fully informative markers located 25 cM from the trait locus. Thus the efficiency of flanking markers, each 25 cM from the trait locus, appears to be roughly comparable to that of a single marker linked at half the distance.

It is interesting that the ELOD for the parametric multipoint analysis was about 10%–20% higher for the mixed and two-locus models than for the "correct" model of independent normal residual variation. This result is not only counterintuitive but appears to conflict with that found in the study by Boehnke (1990), where the mixed model produced lower ELODS than did the random environmental model. Boehnke (1990) showed that, in general, the sample size required to achieve a lod score of 3.0 was 1.0–2.2 times greater under the polygenic model, depending on the degree of separation of genotypes and on the pedigree and sampling strategy used. We note, however, that the cases analyzed by Boehnke were for a rare (q = .025) dominant trait which was completely linked to the marker locus (θ = 0). In these cases the trait locus accounted for relatively small proportions of the total phenotypic variance; for the largest effect examined, the major locus accounted for only 18% of the total variance. It is noteworthy that Boehnke found that the larger the major-gene effect, the less the effect of polygenic background. Thus the large major-locus effect and the high gene frequency used in our study may account for the disparate findings. The MIM approach clearly performed better in the presence of polygenic or additional single-locus variation. In fact, any effect which increases the overall average similarity of the sibship trait values allows the method to better detect the effects of trait loci located in the marker region.

Table 3 shows that, when the parameters of the major locus were estimated from the data, there was a reduction in average lod score for the multipoint linkage method. This was certainly the case for models II and III, in which additional variation was present in the form either of polygenic effects or of a second major locus which could not be accounted for by the model used to estimate the major-locus parameters. However, there was also a smaller reduction in power for the case in which the basic model was correct. As table 3*A* shows, the residual variation was underestimated in models II and III; all the genetic variation tended to be subsumed under the single major locus. The degree of dominance and gene frequency estimates did not seem to be as affected. For the MIM method, it has been our experience that incorrect specification of heritability may result in quite biased estimates of P, the proportion of genetic variance due to

loci in the test marker region, but does not result in loss of power or in inflated type I error.

We chose as our baseline model an additive major locus with two equally frequent alleles. This would allow all three genotypes to be represented in relatively high frequency and would give a broad spectrum of mating types, while minimizing the frequency of uninformative matings. The results in table 4 indicate that the larger effect is one of reduced gene frequency rather than the effect of dominance. For traits with gene frequencies substantially lower than the .2 used in these simulations, one presumably would be sampling through probands with extreme trait phenotypes and would thus have a high proportion of informative matings. One advantage of the parametric approach is its ability to "recognize" uninformative matings, i.e., matings in which both parents have a high probability of being homozygous at the major trait locus. The MIM method, which does not use parental data and, moreover, does not consider genotypes at individual trait loci, has no way of distinguishing these parents or uninformative matings. If the overall phenotypic variance due to the linked locus remains constant, a reduced gene frequency at the major locus results in an increased effect of the gene, i.e., increased separation between genotype means. It is this separation which has been shown to be the major factor in the power of the lod-score method in other studies (Demenais et al. 1988; Boehnke 1990). This may explain why with reduced gene frequency there was essentially no change in MIM average lod score, while that for MLINK increased by about 25%. We believe that incorporating into our method a probability of informativeness as a likelihood-weighting factor may improve its performance for situations in which there are a large proportion of uninformative families in the sampled data set; stratifying the analysis by parental mating type at the major trait locus has been shown to increase the power of the Haseman-Elston method (Amos et al. 1989). Alternatively, one could employ various sampling strategies in an effort to ensure a high proportion of informative matings.

In table 5 the results regarding the effect that marker heterozygosity and distance have on the comparison of the two methods, are of interest, largely because of the absence of any consistent pattern. These changes may represent random fluctuation, with the relative power of the two methods unaffected either by distance between flanking markers or by marker heterozygosity, at least in the range of these parameters that is examined in the simulation study.

We originally developed the MIM method to be applied to those situations in which there was no evidence of a single major locus. The method was designed to be analogous, in human data, to the method that Lander and Botstein (1989) developed for lower organisms and that was successfully applied to localizing quantitative trait loci in tomatoes (Paterson et al. 1988). However, the results of these simulations indicate that the MIM procedure may prove useful for analyzing quantitative trait data even in many situations in which there is evidence for major-locus variation. Although, as in any simulation study, the generalization of results beyond the specific models tested is hazardous, several conclusions may be drawn from our study regarding the use of the MIM method for the major-locus case. In particular, the method may prove useful in situations where (*a*) parental data are unknown; (*b*) the effect of the major locus is small relative to the total variance; (*c*) there is reason to believe that the model may be inaccurately specified; and (*d*) when there is evidence of residual genetic or correlated environmental variance after the major locus is accounted for. We would advocate using the MIM procedure as a multipoint screening tool testing a series of intervals or 50–100-cM regions across the genome. When significant variation due to a particular region is detected, the parametric approach could be used to simultaneously estimate the model parameters and more accurately characterize the position of the trait locus.

While more work is necessary to examine further questions of robustness to deviations from normality, effects of interference, and the presence of common environmental effects, we believe that, for the analysis of many complex traits, the MIM method can provide a useful alternative to parametric linkage analysis. A user-friendly computer program for carrying out the MIM analysis on real data sets is currently under development and can be obtained at no cost by writing to the authors.

Acknowledgment

This work was supported in part by grants CA 36362 and CA 48711 from the National Institutes of Health.

References

Amos CI, Elston RC (1989) Robust methods for the detection of genetic linkage for quantitative data from pedigrees. Genet Epidemiol 6:349–360

Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE (1989) A more powerful robust sib-pair test of linkage for quantitative traits. Genet Epidemiol 6:435–449

Boehnke M (1990) Sample-size guidelines for linkage analysis of a dominant locus for a quantitative trait by the method of lod scores. Am J Hum Genet 47:218–227

Clerget-Darpoux F, Bonaiti-Peilie C, Hochez J (1986) Effect of mispecifying genetic parameters in lod score analysis. Biometrics 42:393–333

Demenais F, Lathrop GM, Lalouel JM (1988) Detection of linkage between a quantitative trait and a marker locus by the lod score method: sample size and sampling considerations. Ann Hum Genet 52:237–246

Goldgar DE (1990) Multiple analysis of human quantitative genetic variation. Am J Hum Genet 47:957–967

Goldgar DE, Oniki R (1990) Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 47 [Suppl]: A136

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lange K (1986) The affected sib-pair method using identity by state relations. Am J Hum Genet 39:148–150

Lange K, Spence MA, Frank MB (1976) Application of the lod method to the detection of linkage between a quantitative trait and qualitative marker: a simulation experiment. Am J Hum Genet 28:167–173

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

——— (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

Ott J (1974) Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet 26:588–597

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335: 721–726

Penrose LS (1935) The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. Ann Eugen 6:133–138

Risch N, Claus E, Guiffra L (1989) Linkage and mode of inheritance in complex traits. In: Elston RC, Spence MA, Hodge SE, MacCluer JW (eds) Multipoint mapping and linkage based upon affected pedigree members: Genetic Analysis Workshop 6. Alan R Liss, New York, pp. 183–188

Skolnick MH, Bishop DT, Thomas A (1989) Effect of sequential sampling rules and ascertainment correction on genetic analysis. In: Elston RC, Spence MA, Hodge SE, MacCluer JW (eds) Multipoint mapping and linkage based upon affected pedigree members: Genetic Analysis Workshop 6. Alan R Liss, New York, pp 177–182

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

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

Wilson AF, Elston RC, Tran LD, Siervogel RM (1991) Use of the robust sib-pair method to screen for single-locus, multiple-locus, and pleiotropic effects: application to traits related to hypertension. Am J Hum Genet 48:862–872