Am. J. Hum. Genet. 44:216–224, 1989

Identifying Pedigrees Segregating at a Major Locus for a Quantitative Trait: an Efficient Strategy for Linkage Analysis

Michael Boehnke* and Patricia P. Moll†,‡

Departments of *Biostatistics, †Epidemiology, and ‡Human Genetics, University of Michigan, Ann Arbor

Summary

Having found evidence for segregation at a major locus for a quantitative trait, a logical next step is to identify those pedigrees in which major-locus segregation is occurring. If the quantitative trait is a risk factor for an associated disease, identifying such segregating pedigrees can be important in classifying families by etiology, in risk assessment, and in suggesting treatment modalities. Identifying segregating pedigrees can also be helpful in selecting pedigrees to include in a subsequent linkage study to map the major locus. Here, we describe a strategy to identify pedigrees segregating at a major locus for a quantitative trait. We apply this pedigree selection strategy to simulated data generated under a major-locus or mixed model with a rare dominant allele and sampled according to one of several fixed-structure or sequential sampling designs. We demonstrate that for the situations considered, the pedigree selection strategy is sensitive and specific and that a linkage study based only on the pedigrees classified as segregating extracts essentially all the linkage information in the entire sample of pedigrees. Our results suggest that for large-scale linkage studies involving many genetic markers, the savings from this strategy can be substantial and that, compared with fixed-structure sampling, sequential sampling of pedigrees can greatly improve the efficiency for linkage analysis of a quantitative trait.

Introduction

An important question in genetics is whether a trait of interest is under the influence of a major locus, that is, a single gene with large effect. For simple Mendelian traits, such as cystic fibrosis or Huntington disease, answering this question is relatively straightforward. For familial traits with more complex etiologies, such as hemochromatosis, diabetes, or coronary heart disease, or for associated quantitative traits such as serum iron, fasting plasma glucose, or serum cholesterol concentrations, establishing the existence of a major locus is more difficult.

Having found evidence for segregation at a major locus for a quantitative trait, we may then attempt to map the major locus to a particular chromosome using linkage analysis. A useful intermediate step is to identify the subset of pedigrees in which major-locus segregation is occurring; for a locus with a rare dominant allele, this corresponds to pedigrees in which the dominant allele is present. Only segregating pedigrees can provide linkage information. There is a close analogy here to the case of mapping a simple Mendelian disease. In that case, a disease is identified, pedigrees are ascertained in which the disease is segregating, and a linkage study is carried out on only those segregating pedigrees. The key difference is that establishing major-locus segregation and selecting segregating pedigrees is more difficult for traits with complex etiologies.

In this paper, we describe a statistical strategy to identify segregating pedigrees. This strategy compares the likelihood of each pedigree under a model allowing for a major locus with the likelihood of the pedigree according to the model that results when major-locus variability is excluded. Pedigrees are classified as segregating if their trait values are more likely under the major-locus model than under the model without the major locus.

Received August 2, 1988; revision received September 27, 1988.

Address for correspondence and reprints: Michael Boehnke, Ph.D., Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109.

© 1989 by The American Society of Human Genetics. All rights reserved.

0002-9297/89/4402-0006$02.00

Using computer simulation, we explore the validity of this pedigree selection strategy to classify pedigrees as segregating or not and compare the efficiency of a linkage study based only on the pedigrees classified as segregating with that of a linkage study using the entire sample of pedigrees. These properties of the pedigree selection strategy are compared across several major-locus and mixed models, all having a rare dominant allele, as well as across several pedigree sampling designs.

Our results suggest that (*a*) the pedigree selection strategy is sensitive and specific, particularly if sampled pedigrees include at least 3 generations; (*b*) a linkage study based on the selected pedigrees extracts essentially all the linkage information present in the entire sample, while requiring marker information on only a fraction of the pedigrees; (*c*) linkage analysis of quantitative traits can be successful even with modest-sized samples; and (*d*) compared with fixed-structure sampling, sequential sampling of pedigrees can significantly improve the efficiency of linkage analysis of a quantitative trait.

Material and Methods

Models

For most of this paper, we consider the case of a quantitative trait determined by a major locus with a rare dominant allele, a case henceforward referred to as the dominant major-locus model. This model includes a two-allele major locus and normally distributed individual-specific environmental exposures. The quantitative trait value is the sum of a major-locus mean and a normally distributed environmental term. Thus, the population trait distribution is a mixture of normals. Parameters of the major-locus model include the frequency q of the rare dominant allele A, means μ_{aa} and $\mu_{Aa} = \mu_{AA}$ for the major-locus genotype distributions, and the within-distribution SD σ. The model that results when the gene frequency $q = 0$ we shall call the random model. Under the random model, the trait is normally distributed, and each individual may be thought of as having the same major-locus genotype aa.

In the Discussion, we also consider the mixed genetic model (Elston and Stewart 1971; Morton and MacLean 1974). The mixed model may be considered as an extension of the major-locus model according to which a fraction h^2 of the within-distribution variability is due to the effects of additive polygenes. The model that results when the gene frequency q for the mixed model is set to zero is called the polygenic model (Fisher 1918).

Strategy to Detect Segregating Pedigrees

For the rest of this section and the Results section, consider the case of a quantitative trait determined by a dominant major-locus model. In samples of pedigrees in which a major locus is segregating, there will be some pedigrees in which the rare major-locus allele A is present and others in which it is absent. The pedigrees in which the rare major-locus allele A is absent will have all individuals with the same major-locus genotype aa. In these pedigrees, the quantitative trait can be thought of as following a random model. Thus, the distribution of the quantitative trait has a different etiologic basis in the two types of pedigrees.

To distinguish these two types of pedigrees, we calculate for each pedigree a test statistic and classify the pedigree as segregating or not depending on the value of the test statistic. The rationale for the pedigree test statistic is that it compares the likelihood of the pedigree under the best-fitting model if the rare major-locus allele A is present (so that the pedigree is segregating) with the likelihood of the pedigree under the best-fitting model if the rare major-locus allele A is absent (so that the pedigree is nonsegregating). We define the pedigree test statistic S as

$$S = \log L_{ML}(\hat{\mu}_{aa},\hat{\mu}_{Aa},\hat{q},\hat{\sigma};x) - \log L_R(\hat{\mu}_{aa},\hat{\sigma};x) , \quad (1)$$

where x is the vector of trait values for the pedigree members, L_{ML} and L_R are the dominant major-locus–model likelihood and the random-model likelihood for the pedigree, respectively, and ˆ indicates maximum likelihood estimate (MLE) for the dominant major-locus model.

We classify a pedigree as segregating if the pedigree test statistic $S > 0$, corresponding to greater likelihood under the major-locus model; we classify a pedigree as nonsegregating if $S \leqslant 0$, corresponding to greater or equal likelihood under the random model.

Properties of the Pedigree Selection Strategy

We consider four different properties of our strategy to detect segregating pedigrees. Two of these properties, sensitivity and specificity, are standard epidemiologic measures of the validity of a classification scheme (Fleiss 1981, p. 4). The two other properties, overall and per-person linkage efficiencies, are measures of the relative efficiency of a linkage study for the quantitative trait based only on those pedigrees classified as

segregating compared with a linkage study based on all the sampled pedigrees.

In the context of the present study, sensitivity is estimated by the proportion of segregating pedigrees (pedigrees with at least one carrier of the rare major-locus allele A) correctly classified as segregating on the basis of a pedigree test statistic $S > 0$. Similarly, specificity is estimated by the proportion of nonsegregating pedigrees (pedigrees with no carriers of the rare major-locus allele A) correctly classified as nonsegregating on the basis of a pedigree test statistic $S \leq 0$. Clearly, the higher the sensitivity and specificity, the more valid the pedigree selection strategy.

Since a key reason to identify segregating pedigrees is to use the selected pedigrees in a subsequent linkage analysis, we consider two measures of the efficiency of this strategy. We define the overall linkage efficiency as the expected maximum lod score for the selected pedigrees divided by the expected maximum lod score for the entire sample of pedigrees. We define the per-person linkage efficiency as the expected maximum lod score per person for the selected pedigrees divided by the expected maximum lod score per person for the entire sample of pedigrees. The overall linkage efficiency reflects the absolute proportion of the expected lod score that remains when the pedigrees classified as nonsegregating are excluded from the linkage analysis. The per-person linkage efficiency measures the proportion of the expected per-person maximum lod score that remains. Though these definitions of efficiency differ from the standard statistical definition of the term, they are useful in the current context.

Outline of the Simulation Study

For each sample, we generated quantitative trait values for pedigree members under a major-locus model; linked marker phenotypes were also simulated. For each pedigree, if the trait value of the potential proband (see below) was in the upper 5% tail of the quantitative trait distribution, the pedigree was ascertained. Relatives of the proband were then sampled based on a fixed-structure or sequential sampling design. Simulation and sampling of pedigrees continued until the desired total sample size of 450 individuals was reached. The major-locus model was fit to each simulated sample of pedigrees, pedigree test statistics were calculated using the resulting MLEs, and the properties of the pedigree selection strategy were estimated.

Simulation and Sampling

Simulation and data analysis were carried out for two dominant major-locus models. For these models, the genotype means were separated by $k = 1.5$ or 2.0 within-distribution SDs, and 2.5% of the population was in the upper distribution ($q = .0126$), so that the major locus was responsible for 5.2% or 8.9% of trait variability, respectively.

Genetic marker genotypes were also simulated for each pedigree member. Markers were assumed to be fully informative, that is, so highly polymorphic that all unrelated individuals had different heterozygous genotypes. To that end, every pedigree original was assigned a different heterozygous marker genotype by assuming Hardy-Weinberg and linkage equilibria, and cosegregation of the trait and marker loci was simulated. For most of the simulations, the recombination fraction θ between the trait and marker loci was assumed to be .10.

We considered the four pedigree structures illustrated in figure 1 (Burns 1982; Burns et al. 1984; Boehnke et al. 1988). These structures included a nuclear family of size five, 3-generation pedigrees of sizes nine and 15, and an extended pedigree of size 45. Each pedigree structure included only one potential proband, designated by an arrow, ensuring that ascertainment was single. For each pedigree, if the trait value of the potential proband was in the upper 5% tail of the quantitative trait distribution, the pedigree was ascertained. Ascertainment was solely on the basis of the quantitative trait; marker phenotypes were not considered.

For fixed-structure sampling, each sample comprised multiple pedigrees of one of these four structures. For sequential sampling (Cannings and Thompson 1977),

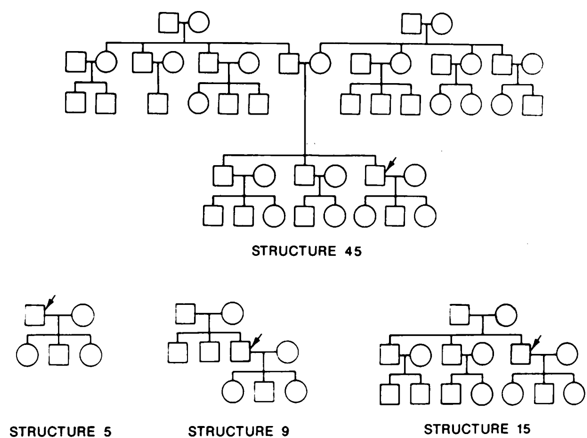

Figure 1 Pedigree structures for the simulated data. Numbers refer to the number of people in each pedigree. Arrows designate the potential probands.

we simulated pedigrees of size 45 and then sampled a subset of each pedigree according to the following nuclear family sequential sampling rule (Thompson 1986; Boehnke et al. 1988): (*a*) If the potential proband has trait value greater than the ascertainment threshold *T*, sample the proband and all first-degree relatives and spouses of the proband. *b*) If any currently sampled individual has trait value greater than *T*, sample any first-degree relatives and spouses of that individual that have not yet been sampled. (*c*) Return to step (*b*) until no further such relatives and spouses remain. (*d*) Continue sampling new pedigrees until the desired total sample size of 450 is obtained. Details of the simulation and sampling process for the quantitative trait are as described by Boehnke et al. (1988).

Calculation of the Pedigree Test Statistics

The simulated samples of pedigrees were analyzed under the dominant major-locus model. Pedigree log likelihoods for this model were calculated using the computer program PAP (Hasstedt and Cartwright 1981), and the sum of these pedigree log likelihoods was maximized to obtain maximum likelihood parameter estimates using the program GEMINI (Lalouel 1979). Since ascertainment was necessarily single (see Simulation and Sampling above), we corrected for ascertainment by conditioning on the quantitative trait value of the proband (Hopper and Mathews 1982; Boehnke 1983; Boehnke and Lange 1984). While this ascertainment correction is not quite as statistically efficient as conditioning on the fact that the trait value of the proband is greater than the ascertainment threshold *T* (Rao et al. 1988), it does correspond to the approach that on both practical and theoretical grounds should almost always be used in the analysis of pedigree data (Young et al. 1988). Pedigree log likelihoods conditional on the proband trait values were then evaluated at both the best-fitting major-locus model and the corresponding random submodel using PAP, with pedigree test statistics calculated as the difference in these pedigree log likelihoods.

Linkage analysis was carried out using the computer program LIPED (Ott 1974, 1976) using the MLEs for the major-locus model. Lod scores were calculated at recombination fractions .01, .05, .10, .15, .20, .25, .30, and .40; maximum lod scores and MLEs of the recombination fraction were then estimated by quadratic interpolation.

Results

Sensitivity and Specificity

Table 1 presents estimates of sensitivity and specificity for the pedigree selection strategy for each combina-

Table 1

Sensitivity and Specificity of the Pedigree Selection Strategy to Detect Segregating Pedigrees

k and Sample Design	Proportion of Pedigrees Actually Segregating	Proportion of Pedigrees Classified as Segregating	Sensitivity	Specificity
1.5:				
5	.225	.347	.710	.759
9	.248	.329	.803	.828
15	.280	.325	.772	.849
45	.417	.422	.746	.810
SS	.242	.290	.800	.873
2.0:				
5	.313	.324	.741	.866
9	.329	.364	.900	.899
15	.360	.360	.858	.921
45	.508	.477	.862	.921
SS	.329	.334	.897	.943

NOTE.—Each set of estimates was based on results pooled over 100 simulated samples of 450 individuals. The data were generated under a dominant major-locus model with genotype means 100 and 100 + 10*k*, within-distribution SD 10, and 2.5% of the population in the upper distribution. Pedigrees were singly ascertained through probands in the upper 5% tail of the distribution. Numbered sample designs are fixed-structure designs based on the pedigree structures shown in fig. 1; SS = sequential sampling based on the 45-person pedigree (see text for sampling designs).

tion of genetic model and sampling design. In each case, these estimates were based on pedigree test statistics pooled over 100 replicate data sets of 450 individuals each. Thus, for each model there were 9,000, 5,000, 3,000, and 1,000 pedigrees of structures 5, 9, 15, and 45, respectively. For the nuclear family sequential sampling rule, there were 4,264 pedigrees for $k = 1.5$ and 3,870 pedigrees for $k = 2.0$. The distribution of the pedigree test statistics for sequential sampling and $k = 2.0$ is given in figure 2.

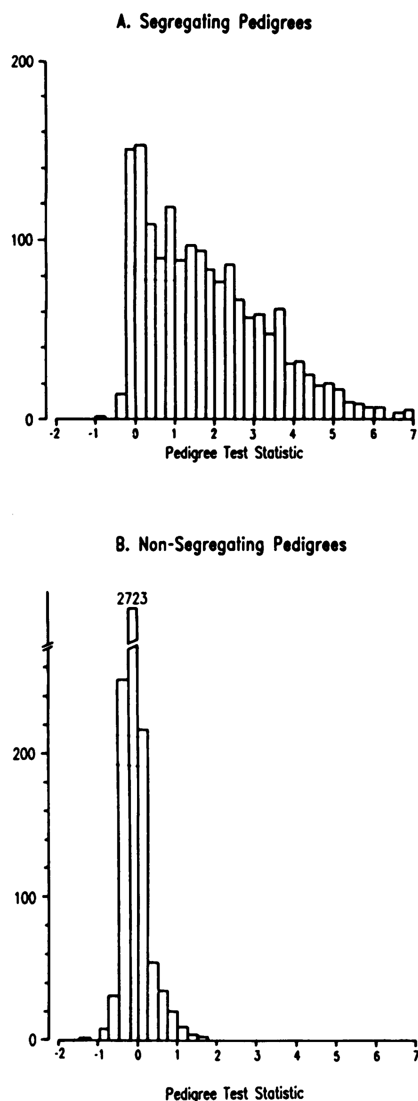

Figure 2 Distribution of the pedigree test statistics. *A*, Distribution for the segregating pedigrees. *B*, Distribution for the non-segregating pedigrees. The data were generated under a dominant major-locus model with genotype means 100 and 120, within-distribution SD 10, and 2.5% of the population in the upper distribution. The pedigrees were singly ascertained through probands in the upper 5% tail of the distribution.

The pedigree selection strategy appears to be sensitive and specific for both models considered. It is not surprising that sensitivity and specificity estimates were greater when the separation between genotype means was greater ($k = 2.0$ vs. $k = 1.5$). Pedigree size and structure appeared to play a more limited role. The proportion of pedigrees classified as segregating tended to be greater than the proportion actually segregating, at least for $k = 1.5$ (table 1).

Linkage Efficiency of the Pedigree Selection Strategy

For the models and sampling designs considered, our pedigree selection strategy proved to be extremely efficient in selecting a subset of the data for subsequent linkage analysis (table 2). In every situation considered, overall linkage efficiency was greater than 90%; in every situation but $k = 1.5$ for pedigree structure 5, overall efficiency was at least 95%. This high overall efficiency was obtained even though the pedigrees selected for linkage analysis typically constituted only one-third of the pedigrees in the sample—except for structure 45, where they constituted just under half the sample (table 1). Sampling variability may have been responsible for the estimated overall linkage efficiency greater than one (structure 15, $k = 1.5$).

The high overall efficiency based on a relatively modest subset of the data yielded impressive per-person linkage efficiencies ranging from 2.08 to 3.10 (table 2). That is, on a per-person basis, a linkage study based only on the selected pedigrees was two to three times as efficient as one based on the entire sample of pedigrees.

This information also allows us to estimate the number of people required to detect linkage to a quantitative trait when the pedigree selection strategy is used (table 2). Dividing the mean maximum lod score for the selected pedigrees by the mean number of individuals selected, then dividing the result into the value 3.0 customarily accepted as conclusive evidence for linkage (Morton 1955), yields the estimates in the final column of table 2. For both models, fixed-structure sampling based on intermediate-sized pedigrees of nine or 15 individuals resulted in greater efficiency for linkage analysis and smaller sample sizes to obtain conclusive evidence for linkage than nuclear families of size five or extended pedigrees of size 45. Sequential sampling using the nuclear family rule was significantly more efficient still: 40% more efficient than the best fixed-structure design (structure 15) for $k = 1.5$ and 39%

Table 2

Efficiency of Linkage Analysis Based on the Pedigree Selection Strategy, Compared with Using All Pedigrees (Recombination Fraction θ = .10)

	Mean Maximum Lod Score		Efficiency		Mean No. of People	Estimated No. of People for a
k and Sample Design	All	Selected	Overall	Per Person	Selected	Lod Score of 3.0
1.5:						
5	.333	.301	.902	2.60	156	1557
9	1.191	1.136	.954	2.90	148	391
15	1.206	1.215	1.007	3.10	146	361
45	1.163	1.145	.984	2.33	190	498
SS	2.275	2.233	.982	2.73	162	218
2.0:						
5	1.307	1.241	.950	2.93	146	353
9	3.512	3.482	.991	2.72	164	142
15	3.411	3.360	.985	2.74	162	145
45	3.754	3.730	.994	2.08	215	173
SS	7.308	7.148	.978	2.16	204	86

Note.—Each set of estimates was based on results pooled over 100 simulated samples of 450 individuals. The data were generated under a dominant major-locus model with genotype means 100 and $100 + 10k$, within-distribution SD 10, and 2.5% of the population in the upper distribution. Pedigrees were singly ascertained through probands in the upper 5% tail of the distribution. Numbered sample designs are fixed-structure designs based on the pedigree structures shown in fig. 1. SS = sequential sampling based on the 45-person pedigree(see text for sampling designs).

more efficient than the best fixed-structure design (structure 9) for $k = 2.0$.

Discussion

Previous Methods to Identify Segregating Pedigrees

Several investigators have previously addressed the issue of identifying pedigrees segregating at a major locus for a quantitative trait. Burns (1982) in a simulation study, Lalouel et al. (1983) in a study of red cell and plasma magnesium concentrations, and Moll et al. (1984) in a study of serum cholesterol levels all used a strategy similar to the pedigree selection strategy described here. With the mixed genetic model as the complete model, they calculated test statistics for each pedigree, on the basis of comparison of the best-fitting mixed model for the entire sample and the best-fitting polygenic model for the entire sample. If the major-locus model rather than the mixed model had been the complete model, their approach would have corresponded to comparing the best-fitting major-locus model and best-fitting random model, by using the pedigree test statistic

$$S' = \log L_{ML}(\hat{\mu}_{aa},\hat{\mu}_{Aa},\hat{q},\hat{\sigma};x) - \log L_{R}(\tilde{\mu}_{aa},\tilde{\sigma};x) , \quad (2)$$

where ˆ and ˜ indicate MLEs under the major-locus and random models, respectively. In contrast, we used the best-fitting major-locus model and the subset of the MLEs for that model that were required by the random model (eq. [1]).

The simulation results of Burns (1982) suggest that the strategy of comparing the best-fitting models is neither sensitive nor specific for identifying segregating pedigrees. She found that the test statistics for the segregating pedigrees were scattered rather uniformly among those for the nonsegregating pedigrees, except at the extreme positive end of the distribution; there, a small excess of segregating pedigrees was found. Burns's results suggest that while their strategy (eq. [2]) is likely to correctly identify a few segregating pedigrees, as in the Moll et al. (1984) study, our pedigree selection strategy (eq. [1]) will provide a much more sensitive and specific means to identify most of the segregating pedigrees in a sample.

We hypothesize that the reason for the greater sensi-

tivity and specificity of our pedigree selection strategy is that it employs parameter estimates that more accurately reflect the two different etiologies present in the samples of segregating and nonsegregating pedigrees. The true model for the segregating pedigrees is the major-locus model with means and SD equal to those that generated the sample. The true model for the nonsegregating pedigrees is the random model with mean and SD equal to the lower mean and SD used to generate the sample. By using (a subset of) the MLEs for the major-locus model (eq. [1]), we come as close to these true models as possible. In contrast, using the best-fitting random model (eq. [2]) results in MLEs for the lower mean and SD that are biased upward from the true values for the nonsegregating pedigrees.

Beaty (1980), Williams and Lalouel (1982), Beaty and Boughman (1986), and Moll et al. (1989) also compared the best-fitting versions of two competing models to identify segregating pedigrees. In each of these studies, the models compared were nonnested; that is, neither model was a submodel of the other. Boehnke (1983) and Boehnke and Lange (1984) used pedigree test statistics devised by Hopper and Mathews (1982) to detect pedigrees inconsistent with the polygenic model, as a means of detecting segregating pedigrees. Their method was neither as sensitive nor as specific as the pedigree selection strategy described here.

Identifying Pedigrees Segregating in the Context of the Mixed Model

Our pedigree selection strategy can be used for any model including a major locus. For the mixed model, one could calculate pedigree test statistics comparing the best-fitting mixed model and the polygenic submodel using the appropriate mixed-model MLEs.

Indeed, the mixed-model case is of particular importance, given its wide use in human genetics. The computational complexity of the mixed model (Hasstedt 1982), particularly in the linkage analysis, led us to choose the major-locus model for our large-scale simulations. However, we also carried out less extensive simulations for the mixed model. For these simulations, we used a modified version of the $k = 2.0$ model, one in which half the variability among individuals with the same major-locus genotype was due to additive polygenes ($h^2 = .50$), and considered two sampling designs: (1) fixed-structure sampling based on structure 9 and (2) sequential sampling using the nuclear family rule. We then calculated test statistics for each pedigree on the basis of the best-fitting mixed model, as well as pedigree test statistics based on the best-fitting major-locus model. We used the MLEs for the major-locus model in the subsequent linkage analysis because of the computational difficulty of linkage analysis with the mixed model.

The results for the mixed-model data analyzed using pedigree test statistics comparing the major-locus and random models (i.e., assuming $h^2 = .00$ when in fact $h^2 = .50$) were similar to those for the major-locus data. The primary difference was that while sensitivity remained high (.855 and .872 for structure 9 and sequential sampling, respectively), specificity was substantially reduced (.691 and .721, respectively; see table 1 for comparison). Because of the decreased specificity, although overall linkage efficiency remained high (.979 and .969), per-person linkage efficiency was reduced (2.00 and 1.63; see table 2 for comparison). When pedigree test statistics were calculated that compared the mixed and polygenic models, specificity improved to .856 and .834; per-person linkage efficiency should similarly improve if the mixed model were used to carry out the linkage analysis. Even without such improvement, these simulation results demonstrate that, under the situations simulated, our pedigree selection strategy provides an efficient basis for a linkage study of a quantitative trait whose distribution is determined by both a major locus and polygenes.

Modifications and Extensions of the Strategy

Our pedigree selection strategy could be modified in several ways. One possibility would be to change the cut point for classifying a pedigree as segregating from zero to some positive number. This modification initially seems attractive since using zero often results in too many pedigrees being classified as segregating (table 1). Arguing against a change in cut point is that it is not clear a priori what that cut point should be, and it seems likely that any attempt to optimize the choice of cut point would yield a result that is model dependent. Since using zero resulted in good sensitivity and specificity for the models and sampling designs considered, a change seems unnecessary.

An alternative strategy to identify segregating pedigrees would be to classify pedigrees as segregating if at least one individual has a high probability of carrying the rare major-locus allele. This probability may be estimated as $[L(x,\mathrm{AA})+L(x,\mathrm{Aa})]/[L(x,\mathrm{AA})+L(x,\mathrm{Aa})+L(x,\mathrm{aa})]$, where $L(x,g)$ is the likelihood of the phenotypes x of the pedigree members evaluated at the MLEs for the major-locus or mixed genetic model, assuming

the individual in question has major-locus genotype g. However, simulation results for a mixed model with a rare dominant allele (Odenheimer 1985) suggest that this approach has poor sensitivity and specificity for assigning individual genotypes when the separation between major-locus genotypes means is 2.0 or less within-distribution SDs. An additional problem is that there is no obvious numerical definition of high probability of carrying the rare major-locus allele. If this latter problem could be solved, one might combine individual genotype probabilities with pedigree test statistics in some overall strategy. Given the success of our pedigree selection strategy for subsequent linkage analysis, this additional complexity would appear unnecessary for the models considered in the present study. A different decision might be appropriate if, rather than linkage analysis, we were interested in determining disease etiology and appropriate treatment modalities for specific individuals.

Linkage Analysis of a Quantitative Trait

Lange et al. (1976) used computer simulation to demonstrate the feasibility of linkage analysis of a quantitative trait. They also considered the case of a dominant major-locus trait and a linked codominant marker. Our more extensive simulation results confirm their findings and demonstrate that linkage analysis of a quantitative trait can be feasible even with $k = 1.5$ SDs between the major-locus genotype means, given an efficient sampling design. Mean sample sizes to demonstrate linkage of course will be greater with a less informative marker, but smaller for a more tightly linked marker. When our pedigree selection strategy was applied to pedigrees sequentially sampled using the nuclear family rule in which a quantitative trait influenced by a major locus with a rare dominant allele ($k = 2.0$) was cosegregating with a totally linked ($\theta = .00$), fully informative marker, the mean lod score for 100 samples of average size 199 individuals was 14.27; using pedigrees of structure 9 in the same situation yielded a mean lod score of 7.17 for 100 samples of average size 164 (data not shown). Thus, as few as 42 or 70 individuals may yield a lod score of 3.0 in this situation.

An interesting additional observation from our simulation study was that the recombination fraction of .10 was generally overestimated (data not shown). Mean estimates for the selected pedigrees ranged from .115 to .149 for $k = 1.5$ and from .108 to .119 for $k = 2.0$; results for the entire samples of pedigrees were nearly identical. Whether the upward bias in the estimate of the recombination fraction was due to small sample size, the use of parameter estimates from the segregation analysis in the subsequent linkage analysis (Clerget-Darpoux et al. 1986), or some other factor is unclear.

Sequential Sampling and the Detection of Segregation and Linkage

Comparison of the results for the different sampling designs suggests that pedigree sampling design plays an important role in the efficiency of the subsequent linkage analysis (table 2). For the situations considered, fixed-structure sampling designs based on intermediate-sized pedigrees of nine or 15 individuals provided substantial linkage information, much more than was provided by fixed-structure designs based on nuclear families of size five or extended pedigrees of size 45. Pedigrees sampled sequentially using the nuclear family rule provided the most linkage information of all.

These results are completely parallel to our findings (Boehnke et al. 1988) and those of Burns (1982) and Burns et al. (1984) on the effects of sampling design on the power of complex segregation analysis. Thus, for planning a study to detect major-locus segregation, linkage, or both for a quantitative trait influenced by a rare dominant allele, our nuclear family sequential sampling rule appears to be a highly efficient sampling strategy.

Conclusion

The pedigree selection strategy described here provided a sensitive and specific method to identify pedigrees segregating at a rare dominant major locus for a quantitative trait for the cases considered. From the point of view of a subsequent linkage analysis, the selected pedigrees contained nearly all the information, but at a fraction of the effort of considering the entire sample of pedigrees. For large-scale linkage studies involving many genetic markers, the savings in using this strategy can be substantial, and should make linkage studies of quantitative traits much more efficient.

Acknowledgments

We thank D. Timothy Bishop, Kenneth Lange, and Elizabeth Thompson for reading and commenting on an earlier version of the manuscript. We thank Sandra Hasstedt, Jean-Marc Lalouel, and Jurg Ott for developing the computer programs PAP, GEMINI, and LIPED. This research was supported by National Institutes of Health grants GM41440 (M.B.), HL24489 (P.P.M.), and HL39107 (P.P.M.). Computing support for this work was provided by the University of Michigan.

References

Beaty, T. H. 1980. Discriminating among single locus models using small pedigrees. Am. J. Med. Genet. **6**:229–240.

Beaty, T. H., and J. A. Boughman. 1986. Problems in detecting etiological heterogeneity in genetic disease illustrated with retinitis pigmentosa. Am. J. Med. Genet. **24**:493–504.

Boehnke, M. 1983. Advances in pedigree analysis: ascertainment, goodness of fit, and optimization. Ph.D. thesis, University of California, Los Angeles.

Boehnke, M., and K. Lange. 1984. Ascertainment and goodness of fit of variance component models for pedigree data. Pp. 173–192 *in* D. C. Rao, R. C. Elston, L. H. Kuller, M. Feinleib, C. Carter, and R. Havlik, eds. Genetic epidemiology of coronary heart disease: past, present, and future. Alan R. Liss, New York.

Boehnke, M., M. R. Young, and P. P. Moll. 1988. Comparison of sequential and fixed-structure sampling of pedigrees in complex segregation analysis of a quantitative trait. Am. J. Hum. Genet. **43**:336–343.

Burns, T. L. 1982. Sampling considerations for the determination of genetic transmission mechanisms in quantitative traits. Ph.D. thesis. University of Michigan, Ann Arbor.

Burns, T. L., P. P. Moll, and M. A. Schork. 1984. Comparisons of different sampling designs for the determination of genetic transmission mechanisms in quantitative traits. Am. J. Hum. Genet. **36**:1060–1074.

Cannings, C., and E. A. Thompson. 1977. Ascertainment in the sequential sampling of pedigrees. Clin. Genet. **12**:208–212.

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

Elston, R. C., and J. Stewart. 1971. A general model for the genetic analysis of pedigree data. Hum. Hered. **21**:523–542.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. **52**:399–433.

Fleiss, J. L. 1981. Statistical methods for rates and proportions. 2d ed. John Wiley & Sons, New York.

Hasstedt, S. J. 1982. A mixed-model likelihood approximation on large pedigrees. Comput. Biomed. Res. **15**:295–307.

Hasstedt, S. J., and P. E. Cartwright. 1981. PAP: pedigree analysis package. Tech. rep. 13, Department of Medical Biophysics and Computing, University of Utah, Salt Lake City.

Hopper, J. L., and J. D. Mathews. 1982. Extensions to multivariate normal models for pedigree analysis. Ann. Hum. Genet. **46**:373–383.

Lalouel, J.-M. 1979. GEMINI: a computer program for optimization of general nonlinear functions. Tech. rep. 14, Department of Medical Biophysics and Computing, University of Utah, Salt Lake City.

Lalouel, J.-M., P. Darlu, J. G. Henrotte, and D. C. Rao. 1983. Genetic regulation of plasma and red blood cell magnesium concentrations in man. II. Segregation analysis. Am. J. Hum. Genet. **35**:938–950.

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

Moll, P. P., T. D. Berry, W. H. Weidman, R. Ellefson, H. Gordon, and B. A. Kottke. 1984. Detection of genetic heterogeneity among pedigrees through complex segregation analysis: an application to hypercholesterolemia. Am. J. Hum. Genet. **36**:197–211.

Moll, P. P., V. V. Michels, W. H. Weidman, and B. A. Kottke. 1989. Genetic determination of plasma apolipoprotein AI in a population-based sample. Am. J. Hum. Genet. **44**:124–139.

Morton, N. E. 1955. Sequential tests for the detection of linkage. Am. J. Hum. Genet. **7**:277–318.

Morton, N. E., and C. J. MacLean. 1974. Analysis of family resemblance. III. Complex segregation of quantitative traits. Am. J. Hum. Genet. **26**:489–503.

Odenheimer, D. J. 1985. An evaluation of the validity of complex segregation analysis in identifying an individual's genotype at a major locus. Ph.D. thesis, University of Michigan, Ann Arbor.

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.

———. 1976. A computer program for linkage analysis of general human pedigrees. Am. J. Hum. Genet. **28**:528–529.

Rao, D. C., R. Wette, and W. J. Ewens. 1988. Multifactorial analysis of family data ascertained through truncation: a comparative evaluation of two methods of statistical inference. Am. J. Hum. Genet. **42**:506–515.

Thompson, E. A. 1986. Partial and conditional likelihoods in pedigree analysis (abstract). American Statistical Association meetings, Chicago.

Williams, W. R., and Lalouel, J.-M. 1982. Complex segregation analysis of hyperlipidemia in a Seattle sample. Hum. Hered. **32**:24–36.

Young, M. R., M. Boehnke, and P. P. Moll. 1988. Correcting for single ascertainment by truncation for a quantitative trait. Am. J. Hum. Genet. **43**:705–708.