Am. J. Hum. Genet. 52:1102–1110, 1993

Incorrect Specification of Marker Allele Frequencies: Effects on Linkage Analysis

Nelson B. Freimer,* Lodewijk A. Sandkuijl,*† and Sally M. Blower‡

*Neurogenetics Laboratory, Department of Psychiatry, University of California, San Francisco; †Department of Medical Genetics, Erasmus University, Rotterdam; and ‡Department of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley

Summary

Most current linkage analyses make use of highly polymorphic DNA markers. Assigning correct allele frequencies for these markers may be extremely difficult in particular study populations. Designation of erroneous frequencies may result in false-positive evidence for linkage, as well as in failure to correctly exclude linkage. These effects are most pronounced in small pedigrees with key individuals unavailable for typing. The power to correctly detect true linkage does not appear to be greatly affected by inaccurate allele frequencies. Before linkage analyses are performed for specific pedigrees, it is recommended that simulation analyses be performed, followed by uncertainty and sensitivity analyses.

Introduction

Computer simulation methods have been used to evaluate the impact of specification of incorrect genetic parameters in linkage analyses. For example, Weeks et al. (1990) showed how, in an actual data set, the use of several diagnostic schemes and a wide range of penetrance values inflated the maximum lod score. Little attention, however, has been paid to erroneous linkage results that are due to misspecifying marker allele frequencies. The possibility of such errors may be increased in current studies because of the predominant use of multiallelic markers such as VNTRs and simple-sequence repeats (SSRs) (Nakamura et al. 1985; Litt and Luty 1989; Weber and May 1989), for which there may be particular difficulties in accurately determining allele frequencies (Devlin et al. 1991). As discussed here, errors in marker allele frequencies are only problematic if it is necessary to reconstruct genotypes of unavailable individuals. This situation may occur commonly, as linkage studies increasingly focus on complex traits characterized by advanced age at onset (such as Alzheimer disease). For such traits, genome searches are performed on large collections of small pedigrees.

Linkage analysis of complex traits is hindered by uncertainty in estimated genetic parameters, such as expected age at onset or completeness of penetrance (Lander 1988). It has been suggested that the use of conservative strategies for analyses of complex disorders could reduce the effects of incorrect parameter estimates. For example, by considering individuals with definite illness as affected and all other individuals as phenotype unknown, one reduces the problems of age-dependent and incomplete penetrance (Freimer and Reus 1992). However, even under conservative approaches, allele frequency (of trait and marker loci) is a parameter which must be specified. Also, pedigrees which are whittled down to permit analysis of homogenous segments may be more sensitive to misspecified allele frequencies than are larger pedigrees.

There are two possible forms of error in specifying allele frequencies for SSR markers. (1) It may not be technically possible to definitively identify specific alleles in a given family. Knowles et al. (1992) have discussed errors associated with these assignments. (2) It may be difficult to estimate the allele frequencies in a particular population under study; such a population may not be well defined, and there may be significant variation in allele frequency between populations. Recent surveys have documented dramatic variation in allele frequencies for RFLP and VNTR markers, not

Received May 12, 1992; final revision received February 5, 1993.

Address for correspondence and reprints: Nelson B. Freimer, Neurogenetics Laboratory, Department of Psychiatry, Box F-0984, University of California, San Francisco, CA 94143.

0002-9297/93/5206-0011$02.00

only between widely dispersed populations (Bowcock et al. 1991) but even between populations which are in geographic proximity (Kidd et al. 1991). Although no large-scale allele frequency comparisons between populations have yet been reported for SSRs, Edwards et al. (1992) found considerable variation between American racial groups for four of five SSR markers.

There has been very little study regarding the effects of incorrect designation of allele frequency on linkage results, although, as noted above, this parameter must always be specified in linkage analysis. In cases where all founders can be directly genotyped at marker loci, the correctness of marker allele frequency is practically irrelevant. However, linkage analyses are increasingly being performed for genetically complex diseases, which tend to have late onset and thus are characterized by families with founders not available for genotyping. In a linkage study of affective disorder and HLA in two families, Kruger et al. (1982) observed that the most extreme lod scores were obtained when one marker allele was considered much rarer than the others. In an examination of several characteristics of highly polymorphic markers, Ott (1992) showed that strong biases in estimation of the recombination fraction (with associated false-positive evidence for linkage) can occur in a particular type of erroneous designation of marker allele frequencies if alleles are wrongly assumed to have equal frequencies when several individuals are untyped at the marker locus. In the current study, we describe the results of a more general examination of the effects of misspecified allele frequencies on linkage analysis and suggest some practical solutions. Our observations are based on linkage simulations using an example pedigree, in a large and a small configuration, and a hypothetical marker, with varying number of alleles. We examine the effects in fixed and sequential sampling methods and in situations of linkage and absence of linkage.

Methods

Two separate steps are required in linkage simulation studies. First, the generating step involves preparing hypothetical marker data for all people in the pedigree for whom DNA samples are available. Such marker data are generated conditional on the phenotypes of all family members, requiring specific assumptions to be made concerning the mode of inheritance of the trait or disease. In this step it is possible to vary the informativeness of the hypothetical marker and its distance from the disease gene. Second, the analysis step involves a regular linkage analysis using the generated marker data. Assumed values for certain parameters, such as recombination frequency, marker allele frequencies, or penetrance may differ in the two steps. We will describe these two steps of our study separately.

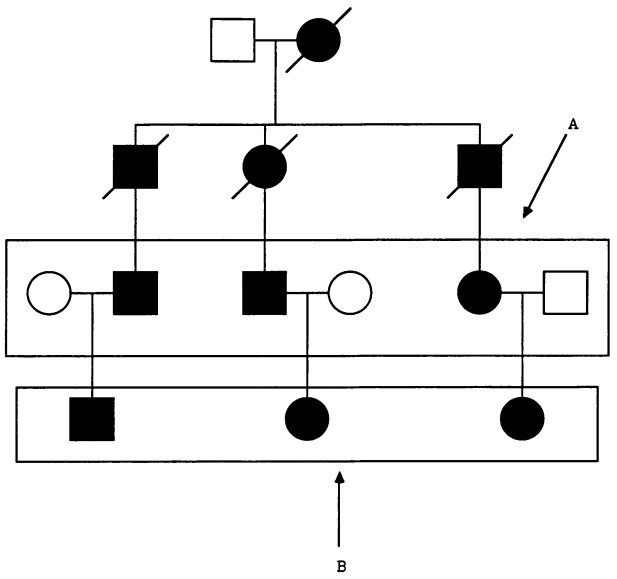

Figure 1 Depiction of pedigree used in simulation analyses. Arrows point to the portions of the pedigree used to construct the L and S configurations. *A*, Individuals available in both the S and L types. *B*, Individuals only available in the L-type configuration.

Pedigree Structure

We performed linkage simulations on an imaginary pedigree in two configurations (designated "L" for large and "S" for small). We constructed a four-generation family tree with only the affected individuals and relevant unaffected spouses in each generation (fig. 1). The L pedigree contained information on nine individuals in generations 3 and 4. The S pedigree, with generation 4 removed, has marker information only on the three affected children in generation 3.

Simulations, Generating Replicates

Marker data were generated via Monte Carlo techniques for two codominant DNA markers with four and eight alleles, by using the MEIOSIM computer program (Sandkuijl and Ott 1989). All simulations assumed an autosomal dominant form of inheritance for the disease, with complete penetrance and a gene frequency of .001. The allele frequencies for the four-allele marker were .4, .2, .2, and .2. For the eight-allele marker, the most common allele had a frequency of .3, with all

others having frequencies of .1. For both markers we generated a series of 1,000 replicates, under the assumption of absence of linkage between the marker and the disease gene. Another series of 1,000 replicates was generated for the four-allele marker, under the assumption of a recombination frequency of 5% with the disease gene. On the basis of our results with the four-allele marker under true linkage, little additional information would have been gained by performing the same analysis with the eight-allele marker.

Simulations, Analysis of Replicates

All analyses were carried out with the MLINK option of the LINKAGE package of computer programs, version 5.04 (Lathrop et al. 1985). For each individual replicate, lod scores were calculated for recombination fractions (θ) increasing from 0 to .5 in steps of .01. All analyses were carried out twice—once with the correct values for the frequencies of the marker alleles and once by assuming that all alleles had equal frequencies ($1/n$). For the disease gene, all parameter values were kept identical to those used in the generating step.

As the L and S configurations both represent relatively small families, we expected that none of the individual replicates would yield a significant lod score. To simulate the practice of actual mapping studies, we therefore combined results from different replicates, just as one would combine results for several families in a fixed-sample approach. We evaluated the effects of pedigree structure and allele frequency in differently sized data sets by varying the number of replicates of the family to be combined in an analysis. This number was varied between 5 and 100. In each case, 10,000 different combinations of the set number of replicates were analyzed. Unlike actual linkage analyses, in our simulations all families had identical pedigree structures. We also simulated a sequential-sample approach by adding replicates until a set lod threshold of +3 was reached.

Using these approaches, we evaluated the possibility of obtaining false-positive, true-positive, and true-negative linkage findings in the various data sets. Results were regarded as evidence for linkage exclusion when lod scores of less than -2 were obtained for values of θ up to .05. When a given data set showed exclusion up to a θ value of .05 and positive evidence for linkage at a greater value of θ, the results were treated as "positive."

Results

Fixed Sampling

Simulation of true linkage.—In the situation of true linkage between the disease and the four-allele marker, supportive evidence was provided by both the L and S pedigrees, with respective average maximum lod scores (per replicate) of 0.662 and 0.283. With allele frequencies for the four-allele marker misspecified as $1/n$ (.25 each), the average maximum lod scores increased marginally to 0.690 for the L pedigree and 0.312 for the S pedigree. The summed lod scores over all 1,000 linked replicates were maximal at .05 recombination frequency (the same value as that used for generating the replicates), when both correct and incorrect marker allele frequencies were used. We combined some of the replicates, chosen at random from the total set of 1,000, into smaller hypothetical data sets (table 1A). When only a few replicates are combined, one cannot expect that the summed lod score for those replicates will always reach its maximum at 5% recombination, nor is it likely that the maximum of that summed lod score can be predicted accurately from the average maximum lod score per replicate, as given above. We evaluated the linkage findings under a wide number of combined replicates (or families). In table 1A we chose to show the number of families, in each scenario, which were required to give a probability of about 90% of finding significant evidence of linkage and also the number which enabled detection of linkage in about 20%–40% of experiments.

When combinations of five replicates of the L type were analyzed, a significant lod score (+3 or more) was obtained in 34.2% of the data sets. With simulated data sets containing 11 families, probabilities of finding significant evidence of linkage increased to about 90%. The gene frequencies specified for the marker alleles had little relevance in this situation (table 1A). As families of the S type provide fewer informative meioses, more families of this type were needed to reach significance; for example, data sets of 11 S-type families only permitted detection of linkage in less than 30% of occurrences. Very few false-negative findings were observed in any of the marker/family configurations (table 1A).

Absence of linkage.—Results were very different for the replicates generated by assuming absence of linkage, especially for the S-type pedigree. With correct allele frequencies, true exclusion of linkage required approximately twice as many families of the S type than of the L type (fig. 2). When incorrect allele frequencies ($1/n$) were used, the probability of correctly excluding linkage was not substantially reduced in the L-type families, except with very small sample sizes (fig. 2 and table 1B.) By contrast, in this situation the number of S-type families required for exclusion was dramatically in-

Table 1

Spurious Linkage and Exclusion Findings for Two Types of Pedigree, Various Degrees of Marker Informativeness, and Various Samples Sizes, Assuming a Fixed Sampling Approach

A. True Linkage Exists (recombination frequency 5%)

No. of Alleles	Allele Frequencies	Pedigree Type[a]	No. of Families[b]	False Exclusion[c]	Detection of Linkage (%)
4	Correct	L	5	10/10,000	34.2
4	Correct	L	11	5/10,000	89.0
4	Incorrect	L	5	6/10,000	36.8
4	Incorrect	L	11	1/10,000	92.3
4	Correct	S	11	2/10,000	21.5
4	Correct	S	29	0/10,000	89.7
4	Incorrect	S	11	0/10,000	29.3
4	Incorrect	S	24	0/10,000	89.9

B. Linkage Exists (recombination frequency 50%)

No. of Alleles	Allele Frequencies	Pedigree Type	No. of Families	Exclusion of Linkage (%)	Spurious Linkage
4	Correct	L	10	91.0	1/10,000
4	Incorrect	L	11	89.6	0/10,000
8	Correct	L	7	90.2	1/10,000
8	Incorrect	L	8	89.3	4/10,000
4	Correct	S	27	89.8	1/10,000
4	Incorrect	S	42	90.0	28/10,000
8	Correct	S	18	90.2	1/10,000
8	Incorrect	S	28	89.9	62/10,000

[a] L = large; and S = small. For further details, see text.

[b] Almost all samples sizes presented here were selected to give a power of approximately 90%, to detect or exclude linkage.

[c] Lod scores ≤ 3 for any value of the recombination frequency were considered as evidence for linkage, otherwise, lod scores ≤ -2 for recombination frequencies $\leq 5\%$ were considered as exclusion of linkage.

creased. Thus, the impact of incorrect allele frequencies on the exclusion power was mostly limited to the S-type pedigree.

The probability of obtaining false evidence for linkage depends on (*a*) the correctness of the marker allele frequencies (and thus, indirectly, on the number of alleles at the marker locus), (*b*) the pedigree structure, (*c*) the number of families in the sample, and (*d*) the threshold for accepting linkage. With correct allele frequencies, the rates of false positives were uniformly low in S and L families, except for a modest increase with very large sample sizes (more than 100 replicates; data not shown). When incorrect allele frequencies were used, a high probability of detecting false evidence for linkage was seen in both the L and S families (table 1B and fig. 3). The false-positive rate was higher in the S families in all scenarios. Unexpectedly, we found that the probability of false positives grew with the size of the data set (fig. 3). With increasing sample size the effect of pedigree structure became progressively more extreme. For example, at the lod threshold of 3, in S-type families with an eight-allele marker, this rate increased from 0.6% in sets of 28 families to more than 4% in sets of 75 families and to 31% for 200 families (or at least 310 times greater than what would be expected). In all scenarios, simulations using the eight-allele marker yielded more false-positive observations than those using the four-allele marker.

The results described above, for *sets* of randomly combined nonlinked replicates, were reflected in the results for *individual* nonlinked replicates. When correct allele frequencies were used, the average lod score for all replicates combined did not become positive for any value of θ, for neither the L nor S pedigrees. With incorrect marker allele frequencies, positive mean lod scores for θ values of .2 or greater were observed

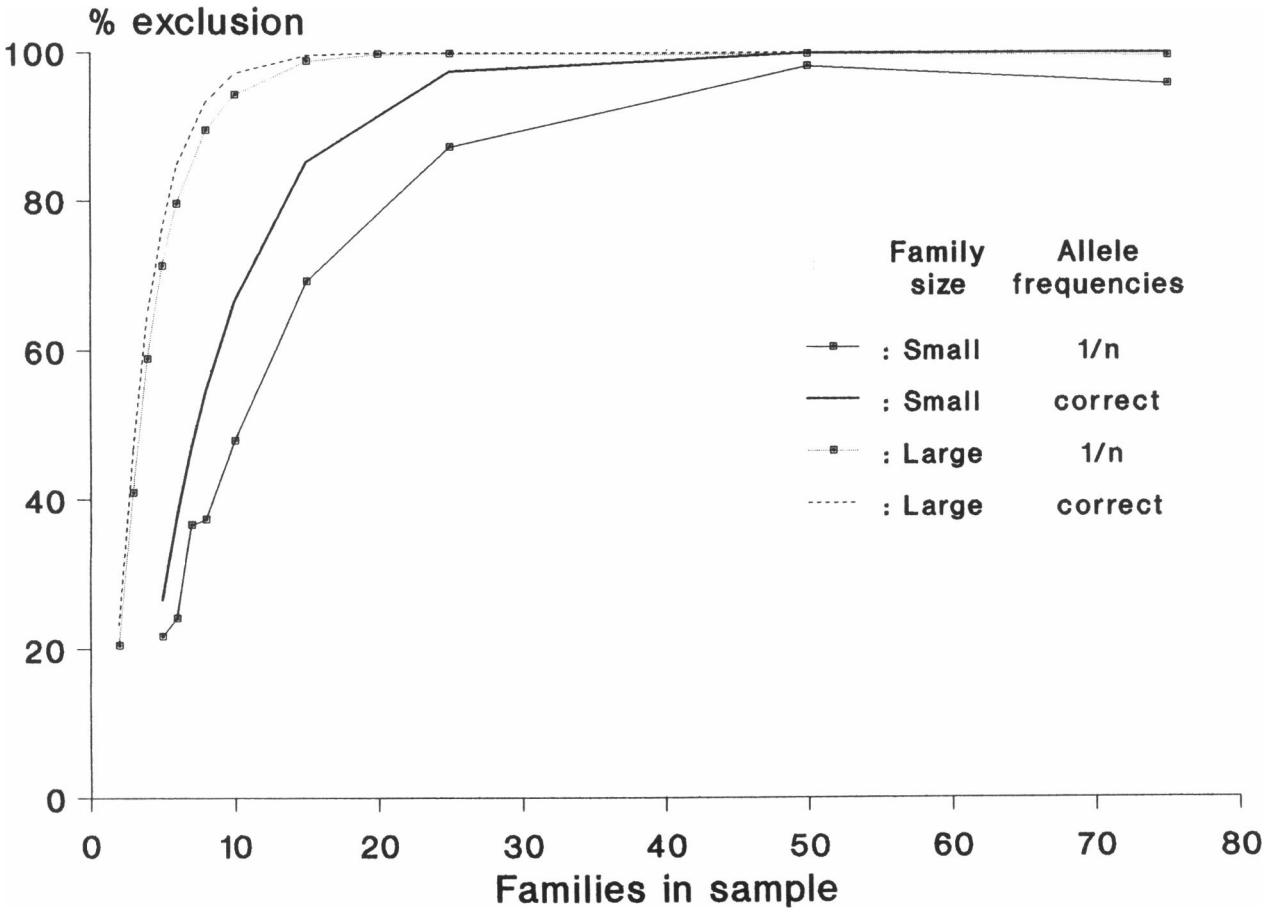

Figure 2 Exclusion power of a marker with eight alleles in the L and S type pedigrees. Results are shown using correct and incorrect allele frequencies and varying the number of families in the sample.

for both S and L types, and for both the four- and eight-allele markers.

Sequential Sampling

In these simulations, (families being added until a set lod score threshold was reached), as in the fixed approach, the use of incorrect allele frequencies led to increased spurious detection of linkage in both types of families, but to a much greater extent in the S families (table 2).

Discussion

1. *Results of Simulation Analyses*

The results of these simulations indicate that incorrect designation of marker allele frequency can have a major effect on linkage analysis, particularly in producing false-positive evidence for linkage. Intuitively, one might predict that using nonextreme values for unknown or uncertain parameters and working with large data sets provides protection against incidental false-positive linkage findings. Our results show that the opposite may be true. We observed higher rates of false positives when we assumed incorrect marker allele frequencies (in this case 1/n) and when we increased the number of families in our analyses. These observations were strongly influenced by the type of pedigree in the data set but were consistent for two different types of study design—the fixed-sample approach and the sequential-sampling approach.

The fixed-sample approach corresponds with the most common way linkage analyses are currently performed, i.e., by simultaneously searching the genome for linkage (lod score greater than or equal to +3) using several hundred highly polymorphic markers on a set collection of families. In some studies, in particular in

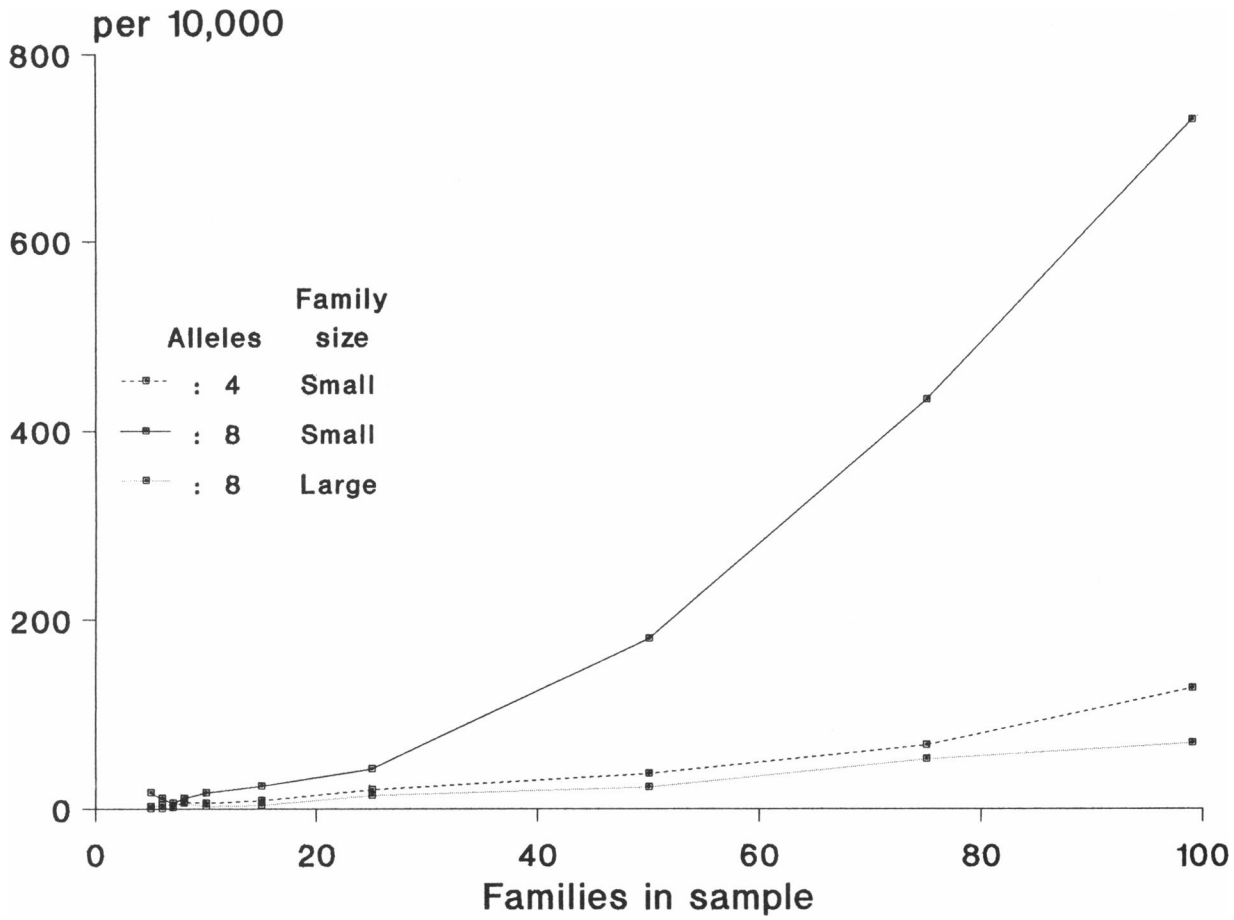

Figure 3 Number of false-positive linkage findings per 10,000 experiments, for DNA markers with four and eight alleles. The effects of increasing the number of families in the sample are shown for both the L and S types of pedigrees. All analyses were carried out under the (incorrect) assumption of equal frequencies for all marker alleles.

collaborative mapping efforts, data from different centers are pooled only for markers that give promising lod scores in one of the participating centers. This latter approach corresponds more closely to sequential sampling, i.e., testing additional families until linkage for a particular marker has definitely been proved or excluded.

A lod score of +3 has been viewed as a stringent criterion, especially when thought of, in simplified terms, as 1,000:1 odds for linkage. In fixed sampling, the asymptotic significance level corresponding to a lod score of +3 is equal to .0001 (one sided) (Ott 1991). The upper bound of this significance level is .001 (Morton 1978), implying that the a priori probability of detecting a false-positive finding is, at worst, 1/1,000. For sequential sampling, the expected rate of false positives is similar to the upper bound of the significance level in fixed sampling (i.e., 1/1,000 at a lod score of +3). Thus, the frequency with which we falsely detected linkage using incorrect allele frequencies (up to 4% for a fixed set of 75 small families, and up to 2.3% in sequential sampling of small families) was greatly in excess of what would be expected. In general, the possibility of drawing erroneous conclusions was greater with increased marker allele number and with small pedigrees that do not permit direct assessment of ancestral marker genotypes. Our results suggest that use of incorrect allele frequencies does not generally lead to failure to detect true linkage, except with very small sample sizes.

The dependence of our findings on different types of pedigree structure was striking. The family types used in this study represent situations in which (for L) significant amounts of linkage information could be obtained in a few such families or (for S) in which many families

Table 2

Spurious Linkage, Assuming Sequential Sampling Approach[a]

No. of Alleles	Allele Frequencies	Pedigree Type[b]	Exclusion of Linkage (%)	Spurious Linkage
4	Correct	L	99.8	6/10,000
4	Incorrect	L	99.4	8/10,000
8	Correct	L	99.2	6/10,000
8	Incorrect	L	98.2	9/10,000
4	Correct	S	99.5	10/10,000
4	Incorrect	S	96.7	114/10,000
8	Correct	S	99.8	16/10,000
8	Incorrect	S	97.5	230/10,000

[a] Sequential sampling was applied with the following upper limits to the total number of families in the sample: small (4 alleles), 50; small (8 alleles), 40; large (4 alleles), 20; and large (8 alleles),16.

[b] L = large; and S = small. For further details, see text.

must be combined to attain significant lod scores. In each case, linkage information is derived by reconstructing genotypes for ancestors not available for direct study. When a given marker allele is shared by all affected descendants, the implication may be that this allele was coinherited with the disease gene (i.e., the alleles are identical by descent) or that several copies of the allele occurred in the top generations (i.e., the alleles are identical by state). Which explanation is favored depends largely on the true population frequency of the shared allele. When the frequency of that allele is underestimated, the probability of identity by descent, due to linkage, will be overestimated. In the L pedigree, this supposed linkage will either be contradicted or confirmed by the offspring in the last generation, while in the S pedigree it cannot be confirmed as to whether the disease is passed to the next generation together with the shared marker allele. False findings are more likely in the S pedigrees because there is less opportunity to directly observe recombinants and nonrecombinants; if only three affected people are available for testing, the highest lod scores are obtained when all three share a rare allele. Families with a structure similar to our L pedigrees are clearly preferable for linkage studies. However, it is likely that in linkage studies of complex traits the pool of available families will include a predominance of S-type and only a few L-type families. For example, this mix characterizes the data sets that have been used in linkage studies of Alzheimer disease (AD) (Schellenberg et al. 1991), and it has been suggested that inaccurate designation of allele frequencies may have been partly responsible for apparent linkage of AD on chromosome 21 which has subsequently been shown not to harbor the major AD locus (Van Broeckhoven et al. 1992). The predominance of pedigrees with an S-type structure is also likely for other disorders with a late onset, such as familial cancers and many psychiatric syndromes. It is worth noting that the seemingly extreme sample sizes (100 or more) which we evaluated here correspond closely to existing collaborative data sets, such as that for tuberous sclerosis (Sampson et al. 1992).

In our study, we used 1/n for the incorrect allele frequencies because this is a commonly made assumption for multiallelic markers, when the true frequency is unknown. However, it seems reasonable that our findings should hold for any designation of allele frequency which is incorrect to a similar degree. This presumption can be tested using the uncertainty analysis methods described later.

The practical importance of our observations depends on the frequency distributions of actual markers. For example, if most markers have a frequency approximating 1/n, then the effect of using this "incorrect" designation should be negligible. The SSR markers currently in wide use are characterized by an extremely variable distribution of allele frequencies within Caucasian reference populations, with the majority characterized by neither a single common allele nor 1/n distribution (Valdes et al. 1993). A recent report of a linkage analysis of panic disorder demonstrated, in the case of a single family, the practical importance of this point (Knowles et al. 1992). For a single marker, the lod score obtained using a marker allele frequency of 1/n was dramatically greater than that obtained using published marker allele frequencies (3.08 compared with 1.25). As more analyses are performed in small incomplete pedigrees, by using SSR markers from across the genome, it is possible that false-positive results will be obtained unless the frequency of each allele can be accurately estimated in the study population. Also, there may be dangers in combining marker data from several sites unless the correctness of allele frequencies has been ascertained for each population. When the effects of inaccurate allele frequencies for pooled data sets are considered, it may be tempting to think that any errors would cancel out because of differing allele frequencies between study populations. In actuality, the use of a single set of allele frequencies in pooled data sets could lead to a situation where incorrect allele frequencies will be used for almost every marker allele in almost every family. As has been shown in our examples, this pooling will not lead to random errors that might can-

cel out but rather to errors that systematically provide evidence for spurious linkage.

2. *Suggested Remedies for the Problem*

In this paper, we used a scenario-based approach to illustrate the potential impact, of misspecifying marker allele frequencies, on estimation of the lod score. As in most linkage simulation studies, we used a fixed-value approach. That is, we varied only one of several input parameters (marker allele frequency) and left all others fixed. This approach is appropriate for exploring the behavior of a specific parameter, but it is too simplistic for analysis of actual data in which uncertainty exists concerning several possibly related parameters. When one parameter is varied at a time, only a small region of the parameter space can be explored, and important effects may be missed. It is possible to more systematically evaluate the magnitude of such effects with techniques that have not been widely used in statistical genetics, such as uncertainty and sensitivity analyses. These methods permit simultaneous variation of the values of all the input parameters and have been applied to explore the behavior of complex models in a wide number of disciplines. They could be used to estimate the effect on the lod score of misspecifying genetic parameters. The variability (or imprecision) in the lod score that is due to the parameter-estimation uncertainty in marker allele frequencies may be determined by an uncertainty analysis. The uncertainty analysis can then be coupled with a sensitivity analysis to determine how the change in value of the lod score is due to the change in the values of the marker allele frequencies. These analyses will identify the magnitude of misspecification errors on the lod score. The combination of efficient sampling strategies for uncertainty analyses with calculation of partial rank correlation coefficients (PRCCs) for sensitivity analyses has been described for epidemiology models (Blower et al. 1991). Use of such a design (*a*) would ensure that the complete range of each parameter is sampled efficiently and without bias and that a frequency distribution for the lod score can be derived and (*b*) would permit qualitative and quantitative evaluation of the independent and the relative effects of the estimation uncertainty, in each of the marker allele frequencies, on the lod score. These analyses can also be extended and coupled with stepwise or rank regression to determine how much of the variation in the lod score is due to each of the key input variables that were identified by their PRCC (Iman et al. 1981*a*, 1981*b*). These methods could be used in uncertainty and sensitivity analyses of single, relatively straightforward parameters, such as marker allele frequencies, but will be most valuable in evaluating interactions of complicated parameters, such as penetrance, age at onset, or disease allele frequency. In a future study, we intend to evaluate these techniques in linkage simulation analyses.

Uncertainty and sensitivity analyses should be performed in advance of genotyping to evaluate the degree to which a particular data set is likely to be vulnerable to spurious linkage findings on the basis of misspecified allele frequencies. Also, when one finds evidence for linkage in such a data set, it is advisable to perform additional analyses by using the actual genotyping data for the "linked" marker(s) and by assuming a wide range of allele frequencies. Use of such a procedure would provide an additional safeguard against false-positive linkage results.

Acknowledgments

We thank Dr. A. Di Rienzo for helpful comments on this manuscript. This work was supported by awards from NARSAD (to N.B.F.), the Lucille Markey Trust (to UCSF), the Simon Fund (from UCSF), and the NIMH.

References

Blower SM, Hartel D, Dowlatabadi H, Anderson RM, May RM (1991) Sex, drugs & HIV: a mathematical model for New York City. Phil Trans R Soc Lond [Biol] 331:171–187

Bowcock AM, Hebert JM, Mountain JL, Kidd JR, Rogers J, Kidd KK, Cavalli-Sforza LL (1991). Study of an additional 58 DNA markers in five human populations from four continents. Gene Geogr 5:151–173

Devlin B, Risch N, Roeder K (1991) Estimation of allele frequencies for VNTR loci. Am J Hum Genet 48:662–677

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–254

Freimer NB, Reus VI (1992) Genetics of bipolar disorder and schizophrenia. In: Rosenberg R, Prusiner S, DiMauro S, Barchi R, Kunkel L (eds) The molecular and genetic basis of neurological disease. Butterworths, New York

Iman RL, Helton JC, Campbell JE (1981*a*) An approach to sensitivity analysis of computer models: Part I—Introduction, input variable selection and preliminary variable assessment. J Quality Technol 13:174–183

——— (1981*b*) An approach to sensitivity analysis of computer models: Part II—Ranking of input variables, response surface validation, distribution effect and technique synopsis. J Quality Technol 13:232–240

Kidd JR, Black FL, Weiss KM, Balazs I, Kidd KK (1991) Stud-

ies of three Amerindian populations using nuclear DNA polymorphisms. Hum Biol 63:775–794

Knowles JA, Vieland VJ, Gilliam TC (1992) Perils of gene mapping with microsatellite markers. Am J Hum Genet 51:905–909

Kruger SD, Turner WJ, Kidd KK (1982) The effect of requisite assumptions on linkage analyses of manic-depressive illness with HLA. Biol Psychiatry 17:1081–1099

Lander E (1988) Mapping complex genetic traits in humans. In: Davies KE (ed) Genome analysis: a practical approach. IRL Press, Oxford

Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Morton NE (1978) Analysis of crossing over in man. Cytogenet Cell Genet 22:15–36

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, et al (1985) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 217:939–941

Ott J (1991) Analysis of human genetic linkage, 2d ed. Johns Hopkins University Press, Baltimore

——— (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283–290

Sampson JR, Janssen LAJ, Sandkuijl LA (1992) Linkage investigation of three putative tuberous sclerosis determining loci on chromosomes 9q, 11q, and 12q. J Med Genet 29:861–866

Sandkuijl LA, Ott J (1989) Determining informativity of marker typing for genetic counseling in a pedigree. Hum Genet 82:159–162

Schellenberg GD, Pericak-Vance MA, Wijsman EM, Moore DK, Gaskell PC Jr, Yamaoka LA, Bebout JL, et al (1991) Linkage analysis of familial Alzheimer disease using chromosome 21 markers. Am J Hum Genet 48:563–583

Valdes AM, Slatkin M, Freimer N (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Van Broeckhoven C, Backhovens H, Cruts M, DeWinter G, Bruyland M, Cras P, Martin J-J (1992) Mapping of a gene predisposing to early-onset Alzheimer's disease to chromosome 14q24.3. Nature Genet 2:335–339

Weber JL, May PE (1989) Abundant classes of human DNA polymorphism which can be typed by the polymerase chain reaction. Am J Hum Genet 44:388–396

Weeks DE, Lehner T, Squires-Wheeler E, Kaufmann C, Ott J (1990) Measuring the inflation of the lod score due to its maximization over model parameter values in human linkage analysis. Genet Epidemiol 7:237–243