Am J Hum Genet 28:167–173, 1976

Application of the Lod Method to the Detection of Linkage Between a Quantitative Trait and a Qualitative Marker: A Simulation Experiment

K. LANGE,[1] M. ANNE SPENCE,[1,2] AND M. BART FRANK[3]

INTRODUCTION

In his extensive review of human linkage studies [1], Renwick questions whether a locus determining a metrical trait can be mapped by statistical analysis of family data. The aim of this report is to dispel some of the pessimism of Renwick's view, at least in a limited number of examples. An assessment is made of the power of the lod method to detect linkage between a quantitative trait and a qualitative marker. Family data on the recently established linkage between the Amy_2 and *Duffy* loci are used by adding Gaussian noise to the Amy_2 phenotypes and calculating the lods for various values of the recombination fraction. The resulting lod curves are then compared for several choices of the noise variance.

FAMILY DATA

A total of 61 informative families with 364 individuals comprised our sample. Of these families 57 were taken from the study by Merritt et al. [2], which established linkage between Amy_2 and *Duffy*. The Medical Genetics Laboratory of the University of California at Los Angeles supplied four more informative families with 21 individuals. The allele frequencies used were $Amy_2{}^A = .946$; $Amy_2{}^B = .054$; $Fy^a = .41$; $Fy^b = .59$ [3, 4].

STATISTICAL METHODS

The Amy_2 locus was converted to a quantitative trait by the following method: the dominant phenotype corresponding to genotypes $Amy_2{}^B/Amy_2{}^B$ and $Amy_2{}^B/Amy_2{}^A$ was assumed to have mean 1, and the recessive phenotype corresponding to genotype $Amy_2{}^A/Amy_2{}^A$ was assumed to have mean 0. Then for each individual in the sample, a random number was drawn from a normal distribution with fixed standard deviation σ and mean

Received November 18, 1974; revised July 1, 1975.

This research was supported in part by the University of California at Los Angeles; grants from the National Institutes of Health, Division of Facilities and Resources (RR-3); Mental Retardation Program and Child Psychiatry Program, Neuropsychiatric Institute, UCLA, (MCH-927); Interdisciplinary Training in Mental Retardation (HD-04612); Mental Retardation Research Center, UCLA (HD-00345); Research Training in Mental Retardation (HD-05615); and Developmental Biology in Mental Retardation.

[1] Department of Biomathematics, University of California at Los Angeles, Los Angeles, California.

[2] Department of Psychiatry, University of California at Los Angeles, Los Angeles, California.

[3] College of Letters and Science, University of California at Los Angeles, Los Angeles, California.

1 or 0 according to the phenotype of the individual at his or her Amy_2 locus. Implicit in this procedure is the assumption that common environment and polygenic inheritance can be neglected. The normally distributed random numbers were generated by the Box and Muller method [5].

Once the Amy_2 phenotypes were converted to quantitative measurements, the likelihood of each family was calculated based on the quantitative phenotype at the former Amy_2 locus and on the phenotype at the *Duffy* locus. The theory behind these calculations is clearly explained by Ott [6], whose computer program was used to perform all calculations. This program requires allele frequencies, possible dominance relations for the qualitative marker, and the mean and standard deviation for each genotype of the quantitative trait. It also presupposes Hardy-Weinberg equilibrium at both loci and linkage equilibrium between them.

For male and female recombination fractions θ_m and θ_f respectively, the likelihood of the whole sample

$$L(\theta_m, \theta_f) = \prod_{j=1}^{n} L_j(\theta_m, \theta_f),$$

where $L_j(\theta_m, \theta_f)$ is the likelihood of family j, and there are n families in all. Lod (θ_m, θ_f) then equals $\log_{10}(L(\theta_m, \theta_f)/L(\frac{1}{2}, \frac{1}{2}))$. If lod $(\theta_m, \theta_f) \geqq 3$ for some choice of (θ_m, θ_f), then traditionally linkage has been accepted [7].

Since quantitative traits do not permit exact determination of genotypes, some measure of the overlap among genotypes is useful. For quantitative traits exhibiting dominance such as our simulated Amy_2 data, a simple index is the difference between the dominant and recessive means divided by the common standard deviation around each mean. A second measure is the percent genetic variance introduced by Elston et al. [8]. We can generalize their notion by observing that the variance of a quantitative trait can be expressed as

$$\Sigma_j p_j \sigma_j^2 + \Sigma_j p_j (\mu_j - \mu)^2,$$

where μ_j and σ_j are respectively the mean and standard deviation for genotype j, p_j is the population frequency of genotype j, and

$$\mu = \Sigma_j p_j \mu_j \text{ [9].}$$

This suggests defining the percent genetic variance as

$$\frac{100\Sigma_j p_j (\mu_j - \mu)^2}{\Sigma_j p_j \sigma_j^2 + \Sigma_j p_j (\mu_j - \mu)^2.}$$

Yet a third measure of overlap is the minimum percent misclassification probability. This measure is applied according to Gold et al. [10] except that we classify an individual as belonging to a given genotype whenever the posterior probability of this contingency is greater than the posterior probability for any other genotype. For traits exhibiting dominance, the misclassification probability applies to discrimination between the dominant and recessive phenotypes.

Both percent genetic variance and minimum percent misclassification probability depend on gene frequencies and apply to random individuals from the underlying population. These two measures of overlap between genotypes are more useful for pedigrees with a large proportion of individuals marrying in at random than for those with large sibships. On the other hand, the ratio of the absolute difference in means to the common standard deviation does not depend on gene frequencies and is adapted best to pedigrees with large sibships. The pedigrees included in the present study tend to have small sibships.

RESULTS

Before simulating the quantitative data, it seemed appropriate to test whether $\theta_m = \theta_f$. This was accomplished by computing the maximum likelihood estimate $\hat{\theta}$ of the common recombination fraction $\theta_m = \theta_f$, and also maximum likelihood estimates $\hat{\theta}_m$ and $\hat{\theta}_f$ under the assumption $\theta_m \neq \theta_f$. The likelihood ratio criterion $\chi^2_1 = 2 \log_e (L(\hat{\theta}_m, \hat{\theta}_f)/L(\hat{\theta}, \hat{\theta}))$ should have an approximate χ^2 distribution with 1 df [11]. For our sample $\hat{\theta} = .21$, $\hat{\theta}_m = .16$, $\hat{\theta}_f = .27$, and $\chi^2_1 = .89$. Since this was not significant for the qualitative data, the assumption $\theta_m = \theta_f$ is employed in the sequel.

The lod curves generated in one typical set of simulation runs are plotted in figure 1 for five choices of σ. A minimum of 10 computer runs were executed for

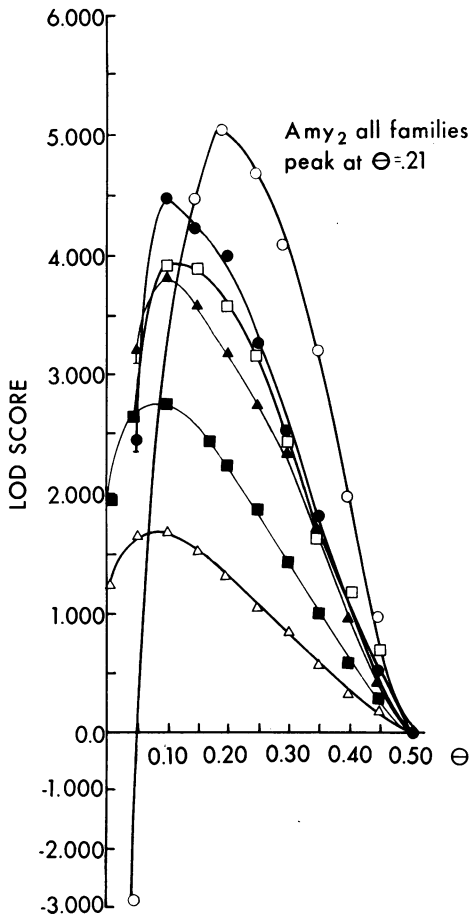

FIG. 1.—Curves for total lod score versus value of recombination fraction θ. *Open circles,* qualitative Amy_2 data; maximum value for $\theta = .21$. Other curves are for data with added Gaussian noise: *closed circles,* $\sigma = .10$; *open squares,* $\sigma = .20$; *closed triangles,* $\sigma = .30$; *closed squares,* $\sigma = .40$; *open triangles,* $\sigma = .50$.

TABLE 1

THREE MEASURES OF GENETIC OVERLAP FOR ACTUAL AND SIMULATED TRAITS

	1	2	3
Amy_2:			
(qualitative)	...	100.0	.0
($\sigma = .1$)	10.0	85.5	.0
($\sigma = .2$)	5.0	59.6	.3
($\sigma = .3$)	3.3	39.6	1.9
($\sigma = .4$)	2.5	26.9	3.7
($\sigma = .5$)	2.0	19.1	4.9
Hypercholesterolemia	3.6	53.7	2.0
Tay-Sachs	...	...	.7
Phenylketonuria	...	...	.4

NOTE.—1 = Absolute difference between distribution means divided by common SD; 2 = percent genetic variance; 3 = minimum percent misclassification probability.

each value of σ. For the sake of comparison, the original qualitative Amy_2 lod curve is also plotted. In table 1, the three measures of overlap are listed for each of the simulated quantitative traits. The same measures are noted for one large pedigree in which hypercholesterolemia is segregating [12]. Minimum percent misclassification figures are given for bivariate measurements on normals and heterozygotes at the Tay-Sachs disease and phenylketonuria loci [10].

DISCUSSION

It is not surprising that θ_m and θ_f do not differ significantly in the present sample. Except for Meyers et al. [13], none of the reported differences [14–17] for other human linkages is statistically significant at the .05 level. For instance, Cook et al. [16] presented a table listing PGM_1-Rh recombinants and nonrecombinants as paternal or maternal. Analyzing this as a two-by-two contingency table does not lead to a significant χ^2 ($P \simeq .14$). The most convincing evidence for sex differences in recombination fractions remains the consistent excess of maternal recombinants over paternal recombinants for many linkages [18].

Several comments about the outcome of the simulation experiment are in order. First, converting Amy_2 to a quantitative trait seems to bias the maximum likelihood estimates of recombination fraction downward. This bias appears consistently in all our computer runs and generally grows more severe as σ increases. We have no explanation for this apparent bias and believe it merits further investigation. Second, it is clear that linkage can still be detected in this "informative" sample when σ rises as high as .30 and perhaps .35. The corresponding measures of genetic overlap offer some guidance about the feasibility of a proposed linkage study. For univariate traits we recommend that all three measures be computed before undertaking linkage analysis. Depending on the nature of the pedigrees available, the three measures can be given different weight as indicated earlier. Table 1 then gives an approximate idea of the limits for each measure

beyond which linkage detection is impractical. Thus, for a hypercholesterolemic pedigree as reported by Elston et al. [12], the amount of genetic overlap is approximately comparable to $\sigma = .30$. For both Tay-Sachs disease and phenylketonuria, the single measure (minimum percent misclassification probability) indicates the possibility of detecting linkage using heterozygotes and normals alone.

Optimism generated by the plots in figure 1 should be tempered with caution. Several parameter values require accurate estimation in an actual study. These include allele frequencies, and perhaps more importantly, means and standard deviations for each genotype. Regression on age and sex may represent another necessary adjustment of the data [8, 12]. If the quantitative trait does not appear to follow a normal distribution around each genotype, it is possible to postulate another parametric family of distributions or to choose an appropriate measurement scale leading to normal distributions. Possible transforms leading to a new measurement scale include the family $x \mid \rightarrow (x^p - 1)/p$ for $p > 0$, with the logarithmic transform as the limiting case $p \rightarrow 0$.

Further hindrances to finding linkages to quantitative traits include the possibility of genetic heterogeneity. One safeguard against this pitfall is to restrict linkage studies to single, large pedigrees. Also, the possibility always exists that no linkage can be detected because all potential marker genes are simply too far from the trait gene. Elston and Lange [19] estimate about a 30% chance that one of 30 markers lies as close to a given trait as *Duffy* does to Amy_2 (i.e. with a recombination fraction of .21 or less). Finally, there is the problem of selecting families informative for the trait locus as well as the marker loci. This problem is avoided in the present study but cannot be avoided in practice. A simple calculation shows that about 19% of all 2 generational families have at least one parent heterozygous at the trait locus Amy_2. More relevant perhaps is the fact that only about 13% of all families having at least one parent heterozygous at the marker locus *Fy* will be informative for linkage. Thus any real sample will contain a large number of uninformative families. It would seem that roughly $61/.13 \simeq 470$ families with at least one parent heterozygous at the marker *Fy* would be necessary to duplicate the amount of information available in our sample.

It would also be desirable to know the multiple x of our sample size which would permit linkage detection for a given σ. A crude estimate of x can be calculated by computing the smallest standard deviation for θ which gives a one-sided confidence interval extending from the true $\theta = .21$ but excluding $\theta = .5$, at the significance level implied by the likelihood ratio criterion lod $(\hat{\theta}) \geqq 3$. According to classical statistical theory [11], the variance of $\hat{\theta}$ is approximately the reciprocal of $-d^2[\log_e L(\theta)]/d\,\theta^2 \mid \theta = \hat{\theta}$. If one fits a parabola to the calculated points of $\log_e (L(\theta)/L(1/2))$, then $d^2[\log_e L(\theta)]/d\theta^2$ is approximately twice the coefficient of θ^2 (i.e., the negative of the curvature of the parabola). This procedure was carried out for all curves corresponding to a given value of σ and the curvatures averaged. For a sample x times as large as our current sample $\hat{\theta} - .21$ will be approximately normally distributed with mean 0 and SD $1/\sqrt{x}$ times the SD

estimated from the average curvature. The smallest value of x giving exclusion of $\theta = .5$ turns out to be .48, .53, .80, 1.53, and 3.11 according as $\sigma = .1$, .2, .3, .4, and .5.

Alternatives to the lod approach outlined above include dichotomizing the quantitative trait by establishing cut off points or thresholds. When the separation between different genotypes is sharp, this causes little harm. With less clean separation, classification by dichotomy can lead to misclassification and occasional genetic inconsistencies within a pedigree. Furthermore, classification by dichotomy is inherently less powerful than an approach based on the full likelihood of the pedigrees considered. The Haseman-Elston sib pair method [20, 21] and the Hill sibship method [22] offer further clever techniques for simple linkage analyses of certain kinds of pedigree data.

SUMMARY

Simulation experiments were performed to assess the power of the lod method for detecting linkage between a quantitative trait and a qualitative marker. Using family data on the *Amy*$_2$*-Duffy* linkage, it was found that linkage detection is feasible in certain limited circumstances. The same qualitative data yielded no evidence for a significant difference in male and female recombination fractions.

ACKNOWLEDGMENTS

We are indebted to Dr. Gary Chase for suggesting that the power of the lod method be checked in the presence of Gaussian noise; to a referee who made useful suggestions for improving the paper; and to Sue Tideman and Mark Johnson, who helped with computer programming. Computer assistance was obtained from the Health Sciences Computing Facility, UCLA.

REFERENCES

1. RENWICK JH: The mapping of human chromosomes. *Annu Rev Genet* 5:81–120, 1971
2. MERRITT AD, LOVRIEN EW, RIVAS ML, CONNEALLY PM: Human amylase loci: genetic linkage with the Duffy blood group locus and assignment to linkage group I. *Am J Hum Genet* 25:523–538, 1973
3. MERRITT AD, RIVAS ML, BIXLER D, NEWELL R: Salivary and pancreatic amylase: electrophoretic characterizations and genetic studies. *Am J Hum Genet* 25:510–522, 1973
4. RACE RR, SANGER R: *Blood Groups in Man,* 5th ed. Oxford, Blackwell, 1968
5. BOX GEP, MULLER ME: A note on the generation of random normal deviates. *Ann Math Stat* 29:610–611, 1958
6. OTT J: Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:588–597, 1974
7. MORTON NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318, 1955
8. ELSTON RC, NAMBOODIRI KK, NINO HV, POLLITZER WS: Studies on blood and urine glucose in Seminole Indians: indications for segregation of a major gene. *Am J Hum Genet* 26:13–34, 1974
9. ELANDT-JOHNSON RC: *Probability Models and Statistical Methods in Genetics.* New York, Wiley, 1971, pp 117–119

10. GOLD RJM, MAAG UR, NEAL JL, SCRIVER CR: The use of biochemical data in screening for mutant alleles and in genetic counselling. *Ann Hum Genet* 37:315–326, 1974
11. RAO CR: *Linear Statistical Inference and its Application.* New York, Wiley, 1973
12. ELSTON RC, NAMBOODIRI KK, GLUECK CJ, FALLAT R, TSANG R, LEUBA V: Study of the genetic transmission of hypercholesterolemia and hyper-triglyceridemia in a 195 member kindred. *Ann Hum Genet.* 39:67–87, 1975
13. MEYERS DA, CONNEALLY PM, HECHT F, LOVRIEN EW, MAGENIS E, MERRITT AD, PALMER CG, RIVAS ML, WANG L: Linkage group I: multipoint mapping, Proceedings 2d International Workshop on Human Gene Mapping. *Cytogenet Cell Genet.* In press, 1975
14. COOK PJL: The Lutheran-secretor recombination fraction in man: a possible sex difference. *Ann Hum Genet* 28:393–401, 1965
15. RENWICK JH, SCHULZE J: Male and female recombination fractions for the nail-patella:ABO linkage in man. *Ann Hum Genet* 28:379–392, 1965
16. COOK PJL, NOADES J, HOPKINSON DA, ROBSON EB: Demonstration of a sex difference in recombination fraction in the loose linkage, Rh and PGM_1. *Ann Hum Genet* 35:239–242, 1972
17. WEITKAMP LR, VAN ROOD JJ, THORSBY E, BIAS W, FOTINO M, LAWLER SD, DAUSSET J, MAYR WR, BODMER J, WARD FE, SEIGNALET J, PAYNE R, KISSMEYER-NIELSEN F, GATTI RA, SACHS JA, LAMM LU: The relation of parental sex and age to recombination in the HLA system. *Hum Hered* 23:197–205, 1973
18. WEITKAMP LR: Human autosomal linkage groups. *Proceedings 4th International Congress Human Genetics,* Paris, September 1971, Amsterdam, Excerpta Medica, 1973, pp 445–460
19. ELSTON RC, LANGE K: The prior probability of autosomal linkage. *Ann Hum Genet* 38:341–350, 1975
20. HASEMAN JK, ELSTON RC: The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3–19, 1972
21. BLACKWELDER WC, ELSTON RC: Comment on Dr. Robertson's communication. *Behav Genet* 4:97–99, 1974
22. HILL, AP: Quantitative linkage: a statistical procedure for its detection and estimation. *Ann Hum Genet* 38:439–449, 1975