Am. J. Hum. Genet. 48:1058–1064, 1991

A Note on Multiple Testing Procedures in Linkage Analysis

Neil Risch

Departments of Epidemiology and Public Health and of Human Genetics, Yale University School of Medicine, New Haven, CT

Summary

Controversy over the impact of multiple testing procedures in linkage analysis is reexamined in this report. Despite some recent claims to the contrary, it is shown that testing multiple markers *decreases* the posterior false-positive rate among significant tests, rather than increasing it; this is true whether the trait of interest is simply monogenic or complex, or even if the genetic model is misspecified. However, if the true mode of inheritance is complex, or if the genetic model is misspecified, the power to obtain a significant result when linkage is present may be reduced, while the significance level is not, leading to an inflation of the posterior false-positive rate. Furthermore, the posterior false-positive rate increases with decreasing sample size and may be unacceptably high for very small samples. By contrast, testing multiple genetic models, by varying either mode-of-inheritance parameters or diagnostic categories, does lead to an inflation of the posterior false-positive rate. A conservative correction for this case is to subtract $\log_{10}t$ from the obtained maximum lod score, where t different genetic and/or diagnostic models have been tested.

Introduction

The lod score method of linkage analysis (Morton 1955) was originally devised as a sequential test of linkage. Among the advantages of the method are (*a*) the ability to include general pedigrees of any form in the analysis and (*b*) the additivity of the statistical evidence (lod score) across pedigrees. The sequential nature of the analysis was assumed to occur through the sequential ascertainment and analysis of pedigrees. On the basis of the theory of Wald (1947), Morton (1955) derived the lod score criterion of 3, with the goal of obtaining a posterior false-positive rate among reported linkages that was <5%. The value of 3 reflects the fact that the prior probability that two randomly selected loci are syntenic and within reasonable mapping distance (say, recombination fraction [θ] < .3) is small, on the order of 2% (Elston and Lange 1975). In fact, linkage analysis is not often performed sequentially; investigators usually have a collection of informative pedigrees which they analyze simultaneously, a procedure which may be more similar to a fixed-sample-size test. Nonetheless, the lod score method and the criterion of 3 have been shown empirically to indeed produce a small false-positive rate among reported linkages (Rao et al. 1978).

As originally described, the statistical criteria were based on the analysis of a single trait-and-marker combination. Since the advent of RFLP and other DNA polymorphism technologies, the number of genetic markers available for linkage analysis in humans has skyrocketed. With the myriad markers now available, concern has been raised that testing for linkage with multiple markers inflates the posterior false-positive rate; that is, given multiple unlinked markers, the probability that at least one of them will produce a (false) positive lod score >3 increases the probability of a false claim of linkage. To remedy this situation, Kidd and Ott (1984) suggested using a lod score criterion of $3 + \log_{10}(g)$, where g is the number of markers tested. Subsequently, Thompson (1984) produced a more precise correction for the number of markers tested, but she basically corroborated the conclusions of Kidd and Ott (1984).

However, Ott (1985) subsequently noted that, when multiple markers are tested, not only does the

Received September 18, 1990; revision received January 15, 1991.

Address for correspondence and reprints: Dr. Neil Risch, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, PO Box 3333, New Haven, CT 06510.

0002-9297/91/4806-0005$02.00

chance of a false lod score of +3 increase, but so does the prior probability of linkage to one of the markers. Ott (1985) showed that, when g markers are tested, the prior probability of linkage increases slightly more rapidly than does the significance level α_g, although the significance level should not be allowed to increase without bound. He therefore concluded that the lod-score-of-3 criterion was still applicable when <100 markers are tested but that if >100 markers are tested the corresponding critical lod score value should be $3 + \log_{10}(g - 100)$.

However, Edwards and Watt (1989) have recently again raised the issue of linkage testing with multiple markers and have concluded that the critical lod score needs to be raised by $\log_{10}(g)$, where g is the number of markers tested, as originally suggested by Kidd and Ott (1984). Furthermore, they have suggested that lack of account of multiple marker testing in the Amish study of affective disorder (Egeland et al. 1987) may have led to a false-positive linkage claim.

Testing of multiple markers is not the only multiple testing procedure that has been employed in linkage studies. For complex diseases, where the true mode of inheritance—and perhaps even the diagnostic boundaries between affected and unaffected—is unknown, multiple analyses may be performed by using different mode-of-inheritance parameterizations and different diagnostic classifications. In the present note, I discuss the issue of when and how it is appropriate to modify the conventional critical lod score value of 3 (or how to interpret obtained lod scores) when multiple testing procedures are employed for various purposes.

Multiple Markers for Mendelian Traits

First consider the case of a simply inherited, homogeneous genetic disease. In this case, a single locus is sought through linkage analysis, and mode-of-inheritance parameters are well defined. The usual procedure is to employ a collection of pedigrees and systematically test a collection of markers from different locations in the genome. It is interesting that such a procedure is typically not sequential with regard to analysis of the pedigrees, which are analyzed simultaneously, but can be viewed as sequential with regard to the markers being tested. When markers give sufficiently negative lod scores (e.g., < -2) with the trait, corresponding regions of the genome are excluded and then other regions are examined. When positive lod scores (>3) occur, the general genome search is stopped, although other nearby markers may be examined to increase the information regarding linkage and the precise location of the trait locus.

For simplification, in the following assume that the markers of interest are both evenly spaced and span the entire genome. A maximum mappable distance of $\theta = .3$ corresponds to a map distance of $w = \sim 34.5$ cM when Kosambi's (1944) mapping function is used. Hence, if each locus spans a distance of $2w = 69$ cM, and if one assumes a total genome length of 3,300 cM (Renwick 1971), then a total of $M = 48$ markers will span the genome. Since the marker loci are 69 cM apart, the θ value between adjacent markers is .44 (by Kosambi's function), so that the marker loci are essentially independent. The prior probability that a trait locus lies within $\theta = .3$ from any marker of interest is $1/48 \approx .02$, similar to the prior probability of linkage for $\theta < .3$, given by Elston and Lange (1975).

Suppose that, for the ith marker locus tested, the probability of getting a significant lod score (>3) when the marker is unlinked to the trait locus, on the basis of its level of polymorphism, is α_i (the significance level), while the probability of getting a significant lod score (>3) when the marker is linked to the trait locus (with $\theta < .3$) is $1 - \beta_i$ (the power). Also suppose that g markers have been tested, with the first $g - 1$ giving nonsignificant lod scores, while the gth marker gives a significant lod score. The question is, What is the probability that this significant test is a false positive, i.e., that the gth marker is actually unlinked to the trait? Let ϕ represent the posterior probability that a significant linkage finding is false (which we call the posterior false-positive rate, as opposed to α, which is the conditional probability of a false-positive result), and let R represent the obtained result. Then

$$\begin{aligned} \phi &= \text{Prob}(\text{unlinked to marker } g \mid R) \\ &= x/(1+x), \end{aligned} \tag{1}$$

where, by Bayes's theorem,

$$x = \frac{\text{Prob}(R \text{ and unlinked to marker } g)}{\text{Prob}(R \text{ and linked to marker } g)} .$$

Now,

$$\begin{aligned} &\text{Prob}(R \text{ and linked to marker } g) \\ &= \frac{1}{M}(1-\beta_g)\prod_{i=1}^{g-1}(1-\alpha_i) , \end{aligned}$$

since the first $g - 1$ tests were not significant and since the trait is unlinked to the first $g - 1$ marker loci. Also,

$$\text{Prob}(R \text{ and unlinked to marker } g) = \sum_{i=1}^{g-1} \text{Prob}(R \text{ and linked to marker } i) + \text{Prob}(R \text{ and unlinked to all markers})$$

$$= \frac{1}{M}\alpha_g \sum_{i=1}^{g-1}(\beta_i)\frac{1}{1-\alpha_i}\prod_{j=1}^{g-1}(1-\alpha_j) + \left(1-\frac{g}{M}\right)\alpha_g \prod_{j=1}^{g-1}(1-\alpha_j)$$

$$= \alpha_g \prod_{j=1}^{g-1}(1-\alpha_j)\left[\frac{1}{M}\sum_{i=1}^{g-1}\beta_i/(1-\alpha_i) + \left(1-\frac{g}{M}\right)\right] .$$

Therefore,

$$x = \frac{\alpha_g}{1-\beta_g}\left[\sum_{i=1}^{g-1}\beta_i/(1-\alpha_i) + (M-g)\right] .$$

Since the α_i are generally quite small, an excellent approximation to the above formula is given by

$$x = \frac{\alpha_g}{1-\beta_g}\left[(M-g) + \sum_{i=1}^{g-1}\beta_i\right]$$

$$= \frac{\alpha_g}{1-\beta_g}\left[(M-1) - (g-1)(1-\bar{\beta})\right] , \qquad (2)$$

where $1 + \bar{\beta}$ is the average power for the first $g - 1$ loci tested. It is obvious from formulas (1) and (2) that $\phi = x/(1+x)$ decreases with increasing g, provided that the α and β values of the markers tested are the same for each marker. The value of x decreases with the $1 - \bar{\beta}$ of all $g - 1$ markers tested so far; it also depends on the specific characteristics of the gth locus, namely, α_g and β_g. Therefore, even though the trend is for x to decrease with g, the value of x for the gth marker may still be high if α_g is large and if $1 - \beta_g$ is small.

The ϕ for practical situations was determined by using formulas (1) and (2) as follows: A fixed sample of N fully informative gametes was assumed, of which R are recombinant. Values of N and R were chosen to give lod scores just above the significance threshold of $+3$. For given values of N and R, the significance level α (the probability of obtaining a lod score $> +3$ when no linkage, or $\theta = .5$, is assumed) was calculated as

$$\alpha = \left(\frac{1}{2}\right)^N \sum_{i=0}^{R}\binom{N}{i} . \qquad (3)$$

$1 = \beta$, which is the probability of obtaining a significant lod score ($> +3$) when linkage with $\theta < .3$ is assumed, was calculated as follows: A uniform distribution for $0 < \theta < .3$ was assumed. Although it is probably more accurate to assume that the map distance, as opposed to the θ value, is uniformly distributed, the close correspondence between the two when $\theta < .3$ suggests that a uniform prior distribution for θ gives an excellent approximation. In fact, the analysis described below was also performed by assuming a uniform prior distribution for the map distance and the Kosambi mapping function relating the θ value and map distance; the results were negligibly different from those described below.

When a uniform distribution for θ on the interval $[0, a]$ is assumed, the probability, in a sample of size N, of a significant lod score ($> +3$), which is obtained with $\leq R$ recombinants, is given by

$$1 - \beta = \sum_{i=0}^{R}\frac{1}{a}\int_0^a \binom{N}{i}\theta^i(1-\theta)^{N-i}d\theta$$

$$= \frac{1}{a}\sum_{i=0}^{R}\binom{N}{i}\int_0^a \theta^i(1-\theta)^{N-i}d\theta$$

$$= \frac{1}{a}\sum_{i=0}^{R}\binom{N}{i}\left\{\frac{1}{N+1}\frac{1}{\binom{N}{i}}\right.$$

$$\left[1-(1-a)^{N+1} - (N+1)a(1-a)^N - \ldots\right.$$

$$\left.\left. - \binom{N+1}{R}a^R(1-a)^{N+1-R}\right]\right\}$$

$$= \frac{1}{a}\frac{1}{N+1}\sum_{i=0}^{R}\left[1 - S(a;N+1,R)\right] , \qquad (4)$$

where $S(a;N+1,R)$ is the sum of the first $R+1$ terms in the binomial expansion of $[(1-a)+a)^{N+1}$.

Formulas (2)–(4) with $a = .3$ were used with various values of N and R to determine α, $1 - \beta$, and ϕ for a single tested marker. The results are given in table 1. Note that, for small N, α is large and $1 - \beta$ is small, so that ϕ is large, as high as 14.5%. In other words, for linkage findings based on small samples, the posterior

Table 1

α, 1 − β, and φ for Single Marker in Fixed Sample of *N* Fully Informative Gametes

N	*R*[a]	α	1 − β	φ
10.....	0	.000977	.297	.145
16.....	1	.000259	.388	.033
20.....	2	.000195	.471	.021
23.....	3	.000237	.548	.022
30.....	5	.000162	.645	.013
39.....	8	.000147	.733	.010
50.....	12	.000153	.821	.010
61.....	16	.000132	.873	.008
71.....	20	.000152	.914	.009
81.....	24	.000159	.941	.009
91.....	28	.000156	.959	.008
101.....	32	.000148	.971	.008

[a] Maximum number of recombinants to obtain a lod score > +3.

probability of their being true is smaller than it is for findings based on larger samples. The φ value appears to stabilize at close to 1% when $N > 30$, at which point α stabilizes at ∼.00015.

To examine the effect of testing multiple markers, a value of $\alpha = .00015$ and a common value of β for all markers was assumed. The φ value when the *g*th marker is the first to show a significant lod score was calculated by using formulas (1) and (2) with $M = 48$. The results are given in table 2. In all cases, the φ value among significant linkages decreases with increasing *g*. The decrease is most rapid, however, when 1 − β is high. It should be noted that, even when all 48 markers have been tested and only the last provides significance, there is a positive probability that this conclusion is false, especially when 1 − β is low. For example, when $1 - \beta = .1$. $\phi = 6.6\%$ for the first marker tested but decreases to only 6.0% when the last (48th) marker is the first one that is significant. This is because there is a high probability that a significant test for a truly linked marker was missed, whereas, when 1 − β is high, missing a true linkage is less likely. The value of 1 − β depends not only on the polymorphism content of the markers but also on the distance between the markers (and hence on the region spanned by each marker). For more closely spaced markers, 1 − β is higher, but also the prior probability of being within the nonoverlapping region of a given marker is lower. Closely spaced markers are not independent, so that consecutive linkage tests with adjacent markers may not be independent but may be positively correlated. For such a case, one would still expect φ to decrease with *g*—but perhaps not as rapidly as indicated in table 2.

Multiple Markers and Complex Traits

A critical assumption in the foregoing discussion is that the trait is due to the same single locus in all pedigrees. If the mode of inheritance is complex, however, this assumption may not hold. For example, for a genetically heterogeneous disease, even with a uniform mode of inheritance, several distinct loci may be involved. Or perhaps the trait is due to a complex interaction among several contributing loci. What effect does such complexity have on the preceding arguments? The essential effect is that the power to detect true linkage (i.e., 1 − β) is decreased, while α may remain the same. The impact, in terms of loss of power, however, will also depend on *N*. For example, if a disease is heterogeneous, and if only a subset is linked to the marker of interest, 1 − β may be drastically reduced when *N* is small, compared with a homo-

Table 2

φ Among Significant Linkages for Sequential Multiple-marker Tests

g	1−β = .9	1−β = .7	1−β = .5	1−β = .3	1−β = .1
1	.008	.010	.014	.023	.066
2	.008	.010	.014	.023	.066
5	.007	.009	.013	.022	.065
10.....	.006	.009	.013	.022	.065
20.....	.006	.007	.011	.020	.063
40.....	.002	.004	.008	.017	.061
48.....	.001	.003	.007	.016	.060

NOTE.—α = .00015 for each marker.

Table 3

$1 - \beta$ and ϕ for Single Marker in Fixed-sized Sample of N Fully Informative Gametes of Which q Are Linked

	$q = .9$		$q = .7$		$q = .5$	
N[a]	$1 - \beta$	ϕ	$1 - \beta$	ϕ	$1 - \beta$	ϕ
10	.187	.212	.069	.422	.023	.687
16	.261	.049	.087	.133	.022	.378
20	.344	.028	.125	.075	.030	.251
23	.428	.028	.178	.064	.045	.214
30	.525	.016	.252	.032	.063	.117
39	.637	.012	.367	.020	.106	.067
50	.740	.011	.495	.016	.175	.043
61	.804	.008	.582	.012	.239	.028
71	.857	.009	.661	.012	.315	.024
81	.893	.009	.719	.011	.382	.021
91	.919	.009	.763	.010	.439	.018
101.....	.938	.008	.798	.009	.489	.015

[a] α Values for each value of N are given in table 1.

geneity situation, but $1 - \beta$ may be reduced only negligibly when N is large.

Specifically, consider the case where only a proportion q of families are due to a particular locus, while the remainder are associated with other causes. If family sizes are small and if all families are combined, the apparent θ will be increased from its true value in linked families to $q\theta + (1-q)/2$ in all families combined. Therefore, for the map of 48 evenly spaced markers as defined above, the span of possible θ values will not be 0–.3 but, rather, from $(1-q)/2$ to $.3q + (1-q)/2$. For this case, $1 - \beta$ can be calculated in a fashion similar to the derivation of formula (4); namely,

$$1 - \beta = \sum_{i=0}^{R} \frac{1}{a} \int_{(1-q)/2}^{aq+(1-q)/2} \binom{N}{i} \theta^i (1-\theta)^{N-i} d\theta$$

$$= \frac{1}{a} \frac{1}{N+1} \sum_{i=0}^{R} \{(S[(1-q)/2; N+1, i]$$

$$- S[aq + (1-q)/2; N+1, i]\} \quad . \qquad (5)$$

Formula (5) with $a = .3$ was used to generate values of $1 - \beta$ and of the corresponding ϕ, for three values of q (= .9, .7, and .5) and for various values of N, the number of fully informative gametes. The value $q = .9$ corresponds to an apparent θ of .05–.32; $q = .7$ corresponds to a range of .15–.36; $q = .5$ corresponds to a range of .25–.40. The results are given in table 3. Two features are quite apparent from the numbers in table 3: $1 - \beta$ and ϕ are quite sensitive to the proportion of linked families (i.e., q) and to N. Under heterogeneity, small samples ($N < 20$) can be disastrous, with unacceptably high ϕ values. With only half of gametes linked, an N of 50 is required to maintain $\phi < 5\%$. The conclusion here, as in the homogeneity case, is that significant linkages based on small N must be viewed cautiously.

Although table 3 was generated under a model of heterogeneity, the principles derived from it apply more generally to other types of genetic complexity. Since it is difficult, if not impossible, to correctly model the mode of inheritance for linkage studies of complex traits, the general trend will be to inflate the apparent θ with linked markers, decreasing $1 - \beta$ and inflating ϕ. Furthermore, multiple contributing loci, epistasis, phenocopies, etc. will decrease the linkage information in a sample, reducing both the effective N and $1 - \beta$. Therefore, when analyzing non-Mendelian traits, one should be particularly conservative in evaluating significant linkage findings based on small N.

Furthermore, in this case as in the monogenic case, ϕ is *never* increased by testing multiple markers but is *decreased,* provided that there is a locus to be found and that $1 - \beta > \alpha$. In the limiting case that no locus exists to be found (i.e., very low $1 - \beta$), all positive results will be equally false, independent of g. Therefore, as pertaining to the comments of Edwards and Watt (1989) regarding the Amish study, a more logical explanation of a false-positive finding of linkage in

that study is that the basic assumption of simple monogenic inheritance as modeled in the analysis is not correct, rather than that multiple markers were tested. In the absence of prior evidence of a monogenic mechanism, it may be prudent to increase the lod score required for significance, to balance the potential effect of a decreased value of $1 - \beta$.

Multiple Models and Diagnostic Classifications

The issue of testing multiple genetic models with different parameterizations and/or different diagnostic classification schemes for affected is not directly analogous to the multiple-marker situation. For this case, assume that t different models and/or diagnostic classifications are examined and that the t lod scores obtained are independent. Although the assumption of independence will not generally hold, the results derived with this assumption will be conservative. For a single marker tested, the maximum lod score across the t different models is used for the significance test. As before, suppose that, for each model tested, the probability of a significant result when linkage is absent is α. Further, suppose that for the ith model the power to detect true linkage is $1 - \beta_i$. In this case, $\phi = x/(1+x)$, where

$$x = \frac{(1-\gamma)[1-(1-\alpha)^t]}{\gamma[1-\sum_{i=1}^{n}(\beta_i)]} \qquad (6)$$

and where $\gamma = 1/M$.

Typically, the $1 - \beta_i$ will be highest for that model (i.e., $_i$) most closely reflecting the true situation, whereas other values of $1 - \beta$ will be much smaller. Hence, if the (nearly) "true" model is included among the t tested, the denominator of formula (6) will be approximately $1 - \beta_i = 1 - \beta$; if the "true" model is not included, the denominator will be close to 0 (i.e., a very high ϕ).

If one assumes that the "true" model is included, formula (6) becomes

$$x = \frac{(1-\gamma)}{\gamma(1-\beta)}[1-(1-\alpha)^t] \ ,$$

which, provided that α is small compared with $1/t$, is well approximated by

$$x = \frac{(1-\gamma)}{\gamma(1-\beta)}\alpha t \ . \qquad (7)$$

Therefore, x, the posterior odds of a false positive, is directly proportional to t. Hence, if one wishes to preserve the same ϕ in this situation, one can use a significance level of $\alpha' = \alpha/t$, which gives the same ϕ as does testing a single (true) model. In terms of lod scores, an $\alpha' = \alpha/t$ can be achieved by using $z + \log_{10}(t)$ for a significance level, or $3 + \log_{10}(t)$ corresponding to the conventional criterion of 3. This correction, originally suggested by Kidd and Ott (1984) for the multiple-marker situation, is conservative. When, in fact, different models produce positively correlated lod scores, the effective value of t would be reduced, and a smaller number than $\log_{10}(t)$ would be added. A more precise approach to this problem has also been described by Ott (1990).

Discussion

Linkage analysis with simple, Mendelian traits has not often been befuddled by confusing and retracted results. Indeed, Rao et al. (1978) showed that, when a lod score criterion of 3 was employed, ϕ for reported linkages was preserved at a low rate, as predicted. Certainly, in these reported linkage studies, multiple markers must have been tested before linkage was found; yet ϕ was not inflated. Although the lod 3 criterion was derived for a sequential test of linkage, it has been shown both here and elsewhere (N. E. Morton, personal communication) that it is robust as a fixed-N test criterion as well.

By contrast, linkage results with several non-Mendelian disorders have led to controversy, with nonreplicable results. As the discussion above suggests, it seems that, rather than concluding that a multiple-marker testing artifact accounts for nonreplicated results for these disorders, it is more reasonable to conclude that the basic assumptions underlying the analysis (monogenic inheritance with known mode of inheritance parameters, one model tested) are incorrect. Although no further protection in lod score analysis is required for the testing of multiple markers, a higher lod threshold may be prudent, to protect against low $1 - \beta$. When multiple models/diagnostic categories are tested, a conservative correction is to subtract $\log_{10}(t)$ for t tests from the obtained loci score. The explanation for the difference in the way multiple testing procedures are handled for the multiple-marker case versus the multiple-model case lies in the fact that, when multiple markers are tested, the prior probability of linkage increases as does the significance level, and so no correction is necessary. With the test-

ing of multiple models, however, the prior probability of linkage does not increase, and the $1 - \beta$, although it may increase to some degree, does not increase linearly—and, in fact, may be near its ceiling with one or a small number of models. Therefore, as indicated by formula (7), the linear increase in the significance level creates an increased ϕ.

In addition, it should be noted that a critical assumption underlying the multiple-model analysis described above is that the models are specified a priori and that new models are not chosen on the basis of observed results from previous models. Such an approach would appear to be beyond statistical correction.

Acknowledgments

I am indebted to N. E. Morton and K. Lange for helpful advice. This work was supported by NIH grant HG00348.

References

Edwards JH, Watt DC (1989) Caution in locating the gene(s) for affective disorder. Psychol Med 19:273–275

Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, et al (1987) Bipolar affective disorders linked to DNA markers on chromosome 11. Nature 325:783–787

Elston RC, Lange K (1975) The prior probability of autosomal linkage. Ann Hum Genet 38:341–350

Kidd KK, Ott J (1984) Power and sample size in linkage studies. Human Gene Mapping 7 (1984): Seventh International Workshop on Human Gene Mapping. Cytogenet Cell Genet 37:510–511

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12:172–175

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

Ott J (1985) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore

——— (1990) Genetic linkage and complex diseases: a comment. Genet Epidemiol 7:35–36

Rao DC, Keats BJB, Morton NE, Yee S, Lew R (1978) Variability of human linkage data. Am J Hum Genet 30: 516–529

Renwick JH (1971) The mapping of human chromosomes. Annu Rev Genet 5:81–120

Thompson EA (1984) Interpretation of LOD scores with a set of marker loci. Genet Epidemiol 1:357–362

Wald A (1947) Sequential analysis. John Wiley, New York