Genetic Epidemiology 12:509–513 (1995)

Recombination Fraction Estimate of Zero in the Presence of Apparent Recombinants: Effects of Incomplete Penetrance and Sporadic Cases

Michael Boehnke and Tempie Hulbert-Shearon

Department of Biostatistics, University of Michigan, Ann Arbor

For a fully penetrant trait, apparent recombinants between the trait and marker loci result in an estimate of the recombination fraction $\theta > 0$. Given allowance for reduced penetrance and/or sporadic cases, this no longer need be true. In this short communication, we describe conditions under which θ is estimated to be zero despite the presence of apparent recombinants. We demonstrate that even if a large proportion of unaffected individuals are apparent recombinants and the penetrance is moderately high, the lod score may be maximized at $\theta = 0$. Despite maximization at $\theta = 0$, presence of apparent recombinants reduces the maximum lod score in comparison to its value if no apparent recombinants are present.

Key words: linkage analysis, lod score, maximum likelihood, recombination fraction, reduced penetrance, sporadic cases

INTRODUCTION

Recently, we carried out a linkage analysis of an autosomal-dominant eye disease segregating in the four-generation pedigree illustrated in Figure 1. In one analysis, we assumed the disease was uncommon, had a penetrance of 0.80, and there were no sporadic cases. For this pedigree, we obtained suggestive evidence of linkage to several markers, including the one illustrated (with minor changes) in Figure 1; for this marker, the maximum lod score of 2.46 was obtained at $\theta = 0$. For fully penetrant genetic traits, such an observation would imply the absence of recombinant individuals in the data. Given reduced penetrance or sporadic cases, this need not be

Received for publication April 17, 1995; revision accepted June 23, 1995.

Address reprint requests to Michael Boehnke, Ph.D., Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029.

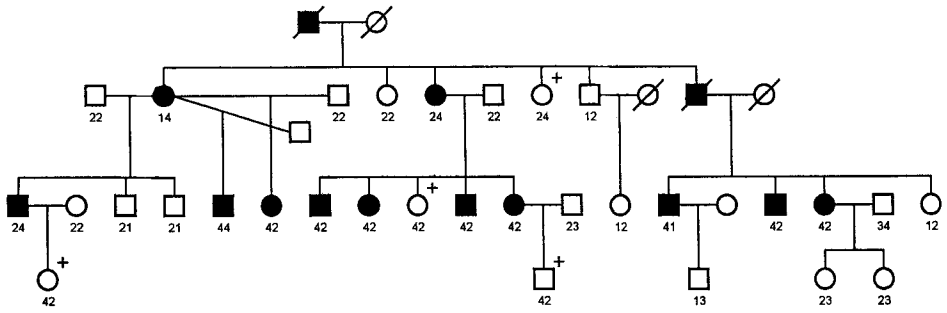

Fig. 1. Autosomal-dominant eye-disease pedigree. Squares are males, circles are females, and slashes indicate deceased individuals. Filled symbols are individuals affected with the disease; open symbols are unaffected individuals. Numbers below each pedigree member provide genotypes at a linked codominant marker. Apparent recombinant individuals are marked by a plus.

the case. Nonetheless, we were interested when in the subsequent haplotyping we identified four apparent recombinants among the unaffected family members (Fig. 1).

Because of this finding, we decided to investigate the proportion of apparent recombinants that might be present in a linkage data set in which the lod score is maximized at zero recombination. We describe this investigation in the context of a dominant disease and offspring of phase-known matings, and evaluate the roles of reduced penetrance and sporadic cases.

METHODS AND RESULTS

Consider the case of an autosomal-dominant disease due to a locus with alleles D and d, with D dominant to d. Let ϕ = P(disease | DD) = P(disease | Dd) be the penetrance and σ = P(disease | dd) be the frequency of sporadic cases, and suppose that these parameter values are fixed in the linkage analysis. Let 1 and 2 be alleles at a codominant marker linked to the disease locus at recombination fraction θ, and suppose that θ is to be estimated. For simplicity, suppose the data are available on offspring of phase-known parental mating type D1/d2 × d1/d1.

In this setting, the disease allele is in coupling with marker allele 1, and offspring may be affected (aff) or unaffected (un), and apparent recombinants (R) or apparent non-recombinants (NR). Conditional on parental mating type, the probabilities of the four types of offspring are

$$\begin{aligned}
\text{P(aff, R)} &= \text{P(aff, 2)} = \tfrac{1}{2}[\theta\phi + (1-\theta)\sigma] \\
\text{P(aff, NR)} &= \text{P(aff, 1)} = \tfrac{1}{2}[\theta\sigma + (1-\theta)\phi] \\
\text{P(un, R)} &= \text{P(un, 1)} = \tfrac{1}{2}[\theta(1-\sigma) + (1-\theta)(1-\phi)] \\
\text{P(un, NR)} &= \text{P(un, 2)} = \tfrac{1}{2}[\theta(1-\phi) + (1-\theta)(1-\sigma)]
\end{aligned}$$

where 1 and 2 indicate the allele transmitted by the affected parent. For example, the probability P(aff, R) of observing an affected, apparent-recombinant offspring equals the probability that the affected parent passes the gamete D2 and that the disease

allele is penetrant $[\theta\phi/2]$, or that the affected parent passes the gamete d2 and that the offspring represents a sporadic case $[(1-\theta)\sigma/2]$.

Let M and N be the number of affected and unaffected offspring, and M_1 and N_1 be the corresponding numbers of apparent recombinants. The natural logarithm of the likelihood of the model is then

$$\begin{aligned}\ln L(\theta;\phi,\sigma) &= M_1 \ln[\theta\phi + (1-\theta)\sigma] + (M - M_1)\ln[\theta\sigma + (1-\theta)\phi] \\ &\quad + N_1 \ln[\theta(1-\sigma) + (1-\theta)(1-\phi)] \\ &\quad + (N - N_1)\ln[\theta(1-\phi) + (1-\theta)(1-\sigma)]. \end{aligned} \tag{1}$$

The parameter values for which $\ln L$ and hence the lod score will be maximized at $\theta = 0$ are those for which the partial derivative of $\ln L$ with respect to θ is negative at $\theta = 0$. To find the boundary on such parameter values, we calculate this partial derivative and evaluate it at $\theta = 0$. A bit of calculus and algebra shows that if $\phi \neq 1$ and $\sigma \neq 0$

$$\begin{aligned}\left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta=0} &= (\phi - \sigma)\left[\frac{M_1}{\sigma} + \frac{M_1 - M}{\phi} + \frac{N_1}{1-\phi} + \frac{N_1 - N}{1-\sigma}\right] \\ &= (\phi - \sigma) f(\phi,\sigma)/[\phi(1-\phi)\sigma(1-\sigma)]\end{aligned}$$

where

$$\begin{aligned}f(\phi,\sigma) &= M_1(\phi + \sigma)(1-\phi)(1-\sigma) - M(1-\phi)\sigma(1-\sigma) \\ &\quad + N_1\phi\sigma(2 - \phi - \sigma) - N\phi(1-\phi)\sigma.\end{aligned}$$

Since $\phi > \sigma$, this partial derivative equals zero if and only if $f(\phi,\sigma) = 0$. Analogous expressions hold for the cases of complete penetrance ($\phi = 1$) and/or no sporadic cases ($\sigma = 0$).

No Apparent Recombinants Among the Affected Offspring

Consider first the case $M_1 = 0$ in which all affected offspring are apparent non-recombinants, as for the eye-disease pedigree. Then $f(\phi,\sigma) = 0$ if

$$\frac{N_1}{N} = \frac{1-\phi}{2-\phi-\sigma}\left[1 + \frac{M(1-\sigma)}{N\phi}\right]. \tag{2}$$

Thus, if the proportion of apparent recombinants among the unaffected offspring is no greater than the term on the right-hand side of Eq. (2), the lod score will be maximized at $\theta = 0$. If in addition the sporadic frequency $\sigma = 0$ and the numbers of affected and unaffected offspring are equal ($M = N$), it turns out that the critical value becomes $(1 - \phi^2)/[\phi(2 - \phi)]$.

Table I presents the maximum proportion of apparent recombinants among the unaffected offspring that result in an estimated θ of zero when all affected offspring are apparent non-recombinants and $M = N$. The proportion depends strongly on the penetrance ϕ, but is nearly independent of the sporadic frequency σ. Obviously, the

TABLE I. Maximum Proportion of Apparent Recombinants Among the Unaffected Offspring for Which the Lod Score Is Maximized at $\theta = 0$ Given Equal Numbers of Affected and Unaffected Offspring and No Apparent Recombinants Among the Affected Offspring

	Sporadic frequency σ		
Penetrance ϕ	$\sigma = 0.0$	$\sigma = 0.1$	$\sigma = 0.2$
1.0	0.00	0.00	0.00
0.9	0.19	0.20	0.20
0.8	0.37	0.38	0.40
0.7	0.56	0.57	0.58
0.6	0.76	0.77	0.77
0.5	1.00	1.00	1.00

lod score cannot be maximized at $\theta = 0$ if apparent recombinants are present among the affected or unaffected individuals when $\phi = 1$, but as ϕ decreases, the proportion that can be apparent recombinants grows rapidly, and reaches 100% for $\phi \leq 0.50$. These large numbers of apparent recombinants that still result in a θ estimate of zero emphasize the substantially greater information provided by the affected offspring than the unaffected offspring when there is reduced penetrance.

While the estimate of θ may not be influenced by presence of some apparent recombinants, the maximum lod score is. Given no apparent recombinants among the affected offspring, and a sporadic frequency $\sigma = 0$, the lod score evaluated at $\theta = 0$ is $(M + N)\log_{10} 2 - N\log_{10}(2 - \phi) + N_1 \log_{10}(1 - \phi)$ [see Eq. (1)]. This lod score decreases linearly with increasing number N_1 of apparent recombinants among the N unaffected offspring (Fig. 2).

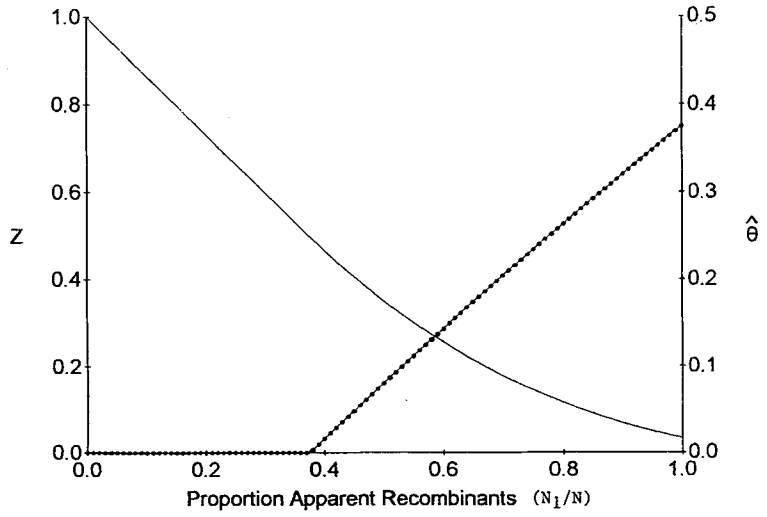

Fig. 2. Effect on the maximum lod score and recombination fraction estimate of proportion of apparent recombinants (N_1/N) among the unaffected offspring. Assumes equal numbers of affected and unaffected offspring, no sporadic cases, and no apparent recombinants among the affected offspring. Z, ratio of maximum lod score given proportion N_1/N apparent recombinants among the unaffected offspring to that when $N_1 = 0$ (solid line). $\hat{\theta}$, maximum likelihood estimate of the recombination fraction (dotted line).

All Offspring Affected

Second, consider the case $N = 0$ in which all offspring are affected. Then $f(\phi, \sigma) = 0$ if the proportion M_1/M of apparent recombinants among the (affected) offspring is $\sigma/(\phi + \sigma)$. In this case, the numbers of apparent recombinants cannot be too large, since generally, the sporadic frequency σ will be substantially smaller than the penetrance ϕ. However, given penetrance $\phi = 0.50$ and sporadic frequency $\sigma = 0.10$, fully $1/6$ of the offspring could be apparent recombinants, and the lod score still would be maximized at $\theta = 0$.

DISCUSSION

For phase-known linkage data, the approach taken here is entirely general, and can be used for any numbers of affected and unaffected offspring. For a rare dominant disease with no sporadic cases, phase-known matings can be distinguished, and our results apply directly. Analogous expressions could be derived for recessive traits. Given a more frequent disease or the presence of sporadic cases, in general it is not possible to distinguish mating types with certainty; furthermore, penetrance and sporadic case frequency are likely to be age dependent. For these more general situations, our results no longer apply directly, but still should be instructive.

The most striking finding of our investigation is the large number of apparent recombinants that may occur among the unaffected offspring and still allow maximization of the lod score at recombination fraction $\theta = 0$. Given equal numbers of affected and unaffected offspring and penetrance $\phi \leq 0.50$, unaffected offspring have no impact on the estimate of θ if the affected offspring all are apparent nonrecombinants (Table I). However, the presence of apparent recombinants, even among the unaffected offspring, does reduce the evidence for linkage as measured by the maximum lod score (Fig. 2).

For the eye-disease pedigree (Fig. 1), approximately $M = 12$ affected and $N = 12$ unaffected offspring were informative. There were $M_1 = 0$ apparent recombinants among the affected and $N_1 = 4$ apparent recombinants among the unaffected. If these offspring all were produced by phase-known matings (which they were not), setting $f(\phi, \sigma) = 0$ would suggest that a penetrance ϕ as high as 0.82 would result in a recombination fraction estimate of zero, consistent with our findings for $\phi = 0.80$.

ACKNOWLEDGMENTS

Support for this work was provided by grant R01-HG00376 (M.B.) from the National Institutes of Health.