*Am J Hum Genet* 35:263–273, 1983

# Linkage between a Marker Locus and a Quantitative Trait of Sibs

C. CLARK COCKERHAM[1] AND B. S. WEIR

SUMMARY

Several variations of a method for detecting linkage between a marker locus and a quantitative trait in full sib families are presented along with computational details. All variations are based on contrasts within qualifying families of three or more sibs. The empirical powers of the various test statistics, evaluated by simulation, were very similar, and also similar to that of Smith. These single-generation tests are likely to be successful only for many families and relatively tight linkage.

## INTRODUCTION

Previously [1], a method of checking for linkage between two characteristics, each of which could be continuously or discretely distributed, was presented. It was essentially an extension of Penrose's [2] graded sib-pair method to include functions of three and four full sibs. With the introduction of comparisons within families of sibs, effects of fortuitous linkage disequilibria and certain environmental correlations could be eliminated, with the restriction, however, that only families of three or more sibs could be utilized. It was concluded that the method had most promise for detecting linkage between a marker locus and a quantitative trait.

The purpose of our study is to translate these theoretical considerations into simplified computational formulas and to provide an approximate test of significance.

## BACKGROUND

We motivate this section by introducing computational formulas of interest. Let the two characteristics be scored as $X$ and $Y$, and distinct sibs within a family be denoted by $i, j, k$, and $l$. The three initial functions of interest for a family of

Received October 26, 1981; revised June 15, 1982.

This study was supported in part by grant GM-11546 from the National Institute of General Medical Services, National Institutes of Health. This is paper no. 8074 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, N.C.

[1] Both authors: Department of Statistics, North Carolina State University, Raleigh, NC 27650.

size $n$ are as follows: average two-sib function,

$$C_2 = \sum_{i \neq j} \sum (X_i - X_j)^2 (Y_i - Y_j)^2 / n(n-1) \ ;$$

average three-sib function,

$$C_3 = \sum_{i \neq j \neq k} \sum \sum (X_i - X_j)^2 (Y_i - Y_k)^2 / n(n-1)(n-2) \ ;$$

average four-sib function,

$$C_4 = \sum_{i \neq j \neq k \neq l} \sum \sum \sum (X_i - X_j)^2 (Y_k - Y_l)^2 / n(n-1)(n-2)(n-3) \ .$$

$C_2$ is Penrose's [2] sib-pair function, and $C_3$ and $C_4$ are extensions of the same function to include three and four sibs, respectively.

Two comparisons within a family are considered: $H_3 = C_2 - C_3$; $H_4 = C_2 - 2C_3 + C_4$; $H_4$ requiring at least four sibs.

Now, let the genotypes at the marker locus be scored as $X$ as follows:

| | $AA$ | $A\bar{A}$ | $\bar{A}\bar{A}$ |
|---|---|---|---|
| Codominant..... | 2 | 1 | 0 |
| Recessive..... | 1 | 0 | 0 |

where allele $A$ has frequency $p$, and $A$, meaning not $A$, has frequency $1 - p$. The expectations, $\mathscr{E}$, of $H_3$ and $H_4$ (given in [1] as $\mathscr{E}H_3 = L_3{}^*$, $\mathscr{E}H_4 = L_4{}^*$) over all families are: $\mathscr{E}H_3 = 5\lambda^2\mathscr{C}^2{}_{ab} + \lambda^2\mathscr{C}_{a2b2} + \lambda^2 k$; $\mathscr{E}H_4 = 4\lambda^2\mathscr{C}^2{}_{ab} + \lambda^2\mathscr{C}_{a2b2} + \lambda^2 k$; where the linkage parameter $\lambda$ is one minus twice the recombination fraction, and is zero with free recombination. Consequently, $H_3$ and $H_4$ are expected to be zero when there is no linkage. While $\mathscr{C}^2{}_{ab}$ and $\mathscr{C}_{a2b2}$ (detailed in [1]) involve linkage disequilibrium between the marker locus and a locus affecting the quantitative trait, they do not contribute to $H_3$ and $H_4$ unless the loci do not recombine freely. The term $k$ involves additive variance, $2\sigma_b^2$, and dominance variance, $\sigma_c^2$, for the quantitative locus; $k = p(1-p)(2\sigma_b^2 + \sigma_c^2)$ for a codominant marker gene; and $k = p^2(1-p^2)(2\sigma_b^2 + \sigma_c^2) + (\lambda^2/4)p^2(1-p)^2\sigma_c^2$ for a recessive marker gene ($\lambda^2 k = 4K$ in [1]).

The computational formula can be simplified considerably by utilizing deviations from family means denoted as $x$ and $y$. Three terms suffice with the following coefficients for each function:

| | $\sum_i x_i^2 y_i^2$ | $\sum_i x_i^2 \sum_i y_i^2$ | $(\sum_i x_i y_i)^2$ |
|---|---|---|---|
| $C_2$..... | $\frac{2}{n-1}$ | $\frac{2}{n(n-1)}$ | $\frac{4}{n(n-1)}$ |
| $C_3$..... | $\frac{1}{n-1}$ | $\frac{3n-2}{n(n-1)(n-2)}$ | $\frac{-4}{n(n-1)(n-2)}$ |
| $C_4$..... | $\frac{-4}{(n-2)(n-3)}$ | $\frac{4(n^2-3n+1)}{n(n-1)(n-2)(n-3)}$ | $\frac{8}{n(n-1)(n-2)(n-3)}$ |
| $H_3$..... | $\frac{1}{n-1}$ | $\frac{-(n+2)}{n(n-1)(n-2)}$ | $\frac{4}{n(n-2)}$ |
| $H_4$..... | $\frac{-4}{(n-2)(n-3)}$ | $\frac{4}{n(n-1)(n-2)(n-3)}$ | $\frac{4}{n(n-3)}$ . |

EXPECTATIONS AND VARIANCES UNDER THE NULL HYPOTHESIS

For these developments under the null hypothesis of $\lambda = 0$, we consider the $x$'s to be known without error and the $y$'s to have mean zero, variances $(n-1)\sigma^2/n$, and covariances $-\sigma^2/n$, where $\sigma^2$ is the variance of $Y$'s within families. Then

$$\mathscr{E} \sum_i x_i^2 y_i^2 = (n-1)(\sum_i x_i^2)\sigma^2/n$$

$$\mathscr{E} \sum_i x_i^2 \sum_i y_i^2 = (n-1)(\sum_i x_i^2)\sigma^2$$

$$\mathscr{E} (\sum_i x_i y_i)^2 = (\sum_i x_i^2)\sigma^2 ,$$

leading to $\mathscr{E} H_3 = \mathscr{E} H_4 = 0$ as expected.

Derivations of the variances $V_3$, $V_4$ of $H_3$, $H_4$ are given in APPENDIX A. They are functions of the central moments $\sigma^4 = [\mathscr{E}(Y - \mathscr{E} Y)^2]^2$, $\mu_4 = \mathscr{E}(Y - \mathscr{E} Y)^4$ within families. Note that $Y$ values are independent within families, and that $\mu_4 = 3\sigma^4$ for a normally distributed trait.

COMBINING INFORMATION OVER FAMILIES

Ordinarily we would weight $H$'s with the reciprocal of their variances in combining over families. We want the weights to increase as family size increases and as $X$ variation within families increases. Only the first of these properties is met by the variances $V_3$ and $V_4$, and a possible solution is to adopt a regression approach. For example, if we were to regress $(y_i - y_j)^2$ on $(x_i - x_j)^2$, the regression coefficient would be

$$B_2 = \sum_{i \neq j} \sum (x_i - x_j)^2 (y_i - y_j)^2 / \sum_{i \neq j} \sum (x_i - x_j)^4 .$$

Hence, $B_2 = SC_2$, where

$$S = n(n-1) / \sum_{i \neq j} \sum (x_i - x_j)^4 , \qquad \sum_{i \neq j} \sum (x_i - x_j)^4 = 2n \sum_i x_i^4 + 6(\sum_i x_i^2)^2 .$$

Proceeding in the same manner for the three- and four-sib functions, $B_3 = SC_3$, $B_4 = SC_4$, so that these new comparisons within families also have simple relations to the previous ones: $H_6 = B_2 - B_3 = SH_3$, $H_8 = B_2 - 2B_3 + B_4 = SH_4$. The corresponding variances are: $V_6 = S^2V_3$, $V_8 = S^2V_4$.

We assume a normally distributed trait in order to compare the various $H$'s. With normality

$$V_3 = V_3^*\sigma^4 , \qquad V_6 = S^2V_3^*\sigma^4 = V_6^*\sigma^4$$

$$V_4 = V_4^*\sigma^4 , \qquad V_8 = S^2V_4^*\sigma^4 = V_8^*\sigma^4 ,$$

where

$$V_3^* = \frac{2[(12n - n^2 - 12)(\sum_i x_i^2)^2 + n(n^2 + 4n - 4)\sum_i x_i^4]}{n^2(n-1)^2(n-2)}$$

$$V_4^* = \frac{32[(n^2 - 3n + 3)(\sum_i x_i^2)^2 - n(n-1)\sum_i x_i^4]}{n^2(n-1)(n-2)(n-3)^2}$$

In table 1, we display the $V^*$'s for families of size 3, 4, and 5, and a representative of every possible marker genotypic array. It can be shown that $H_4 = H_8 = 0$ and that $V_4^* = V_8^* = 0$ when only one $X$ is distinct from the others in a family. Except for possible other idiosyncrasies, it was thought that $H_4$ or $H_8$ would be the preferred comparison in families of four or more because of the symmetry involved. We can see in table 1, however, that this is not the case. While $V_4^*$ does decrease as family size increases, it increases as marker variation within families

TABLE 1

VARIANCES OF THE H'S FOR SELECTED FAMILY SIZES AND MARKER GENOTYPE ARRAYS

| | NO. | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | $X = 2$ | $X = 1$ | $X = 0$ | Var($X$) | $V_3^*$ | $V_6^*$ | $V_4^*$ | $V_8^*$ |
| 3..... | 2 | 1 | 0 | 0.333 | 1.000 | 2.250 | ... | ... |
| | 1 | 1 | 1 | 1.000 | 9.000 | 0.250 | ... | ... |
| | 2 | 0 | 1 | 1.333 | 16.000 | 0.141 | ... | ... |
| 4..... | 3 | 1 | 0 | 0.250 | 0.333 | 1.333 | 0 | 0 |
| | 2 | 2 | 0 | 0.333 | 0.333 | 0.750 | 1.333 | 3.000 |
| | 1 | 2 | 1 | 0.667 | 2.111 | 0.190 | 1.333 | 0.120 |
| | 1 | 1 | 2 | 0.917 | 3.444 | 0.101 | 5.333 | 0.157 |
| | 1 | 0 | 3 | 1.000 | 5.333 | 0.083 | 0 | 0 |
| | 2 | 0 | 2 | 1.333 | 5.333 | 0.047 | 21.333 | 0.188 |
| 5..... | 4 | 1 | 0 | 0.200 | 0.167 | 1.042 | 0 | 0 |
| | 3 | 2 | 0 | 0.300 | 0.170 | 0.472 | 0.320 | 0.889 |
| | 1 | 3 | 1 | 0.500 | 0.837 | 0.173 | 0.320 | 0.066 |
| | 2 | 2 | 1 | 0.700 | 1.290 | 0.089 | 1.173 | 0.081 |
| | 4 | 0 | 1 | 0.800 | 2.667 | 0.065 | 0 | 0 |
| | 3 | 1 | 1 | 0.800 | 1.847 | 0.045 | 1.280 | 0.031 |
| | 2 | 1 | 2 | 1.000 | 1.980 | 0.043 | 3.413 | 0.074 |
| | 3 | 0 | 2 | 1.200 | 2.720 | 0.030 | 5.120 | 0.056 |

increases, and this is not overcome by the regression quantity $V_8$*. The three-sib function $V_3$* has the same properties, but $V_6$* decreases as within-family variation increases. Since $V_6$* still decreases as family size increases, it appears that $V_6$ will provide a desirable weighting function over families. We will not consider $H_4$ or $H_8$ any further.

For any of the $H$ functions, then, we can form either weighted or unweighted sums over families. Normal deviates may be formed, approximately, as these sums divided by their standard deviations. Denoting families by $f$: unweighted test statistics,

$$t_{3u} = \sum_f H_{3f} / \sqrt{\sum_f \hat{V}_{3f}} , \qquad t_{6u} = \sum_f H_{6f} / \sqrt{\sum_f \hat{V}_{6f}} ;$$

weighted test statistics,

$$t_{3w} = \sum_f (H_{3f}/\hat{V}_{3f}) / \sqrt{\sum_f (1/\hat{V}_{3f})} , \qquad t_{6w} = \sum_f (H_{6f}/\hat{V}_{6f}) / \sqrt{\sum_f (1/\hat{V}_{6f})} .$$

In APPENDIX B, we show how $V_{3f}$, $V_{6f}$ may be estimated as $\hat{V}_{3f}$ and $\hat{V}_{6f}$.

These test statistics can be used to test the null hypothesis that $\lambda$ (or $H_3$ or $H_4$) is zero against the one-sided alternative that $\lambda$ (or $H_3$ or $H_4$) is positive.

## SMITH'S TEST

A test with some similarities to the above procedure, in that quantities are measured within families and then summed over families, was given by Smith [3] and discussed further by Bener et al. [4]. In a notation consistent with the present case, the procedure is as follows. If the $n_r$ individuals in a family with an $X$ value of $r$($r$ = 0, 1, 2 for a codominant marker) have $y$ values totaling $S_r$, then

$$n = \sum_r n_r , \qquad 0 = \sum_r S_r ,$$

and we set

$$N_2 = \sum_r n_r^2 , \qquad N_3 = \sum_r n_r^3 , \qquad \Omega = \sum_r S_r^2 .$$

The quantity $\Omega$ is the basis of the test, and under the hypothesis of no linkage, it has the same expectation as

$$\eta = (n - N_2/n) \sum_i y_i^2 / (n - 1)$$

and variance estimated as

$$\hat{V}_S = [N_2 (\sum_i y_i^2)^2 + 3(3 - N_2) \sum_i y_i^4]/3 - \eta^2 , \qquad n = 3$$

$$= [\alpha_2 (\sum_i y_i^2)^2 + \alpha_4 \sum_i y_i^4]/n(n-1)(n-2)(n-3) - \eta^2 , \qquad n \geq 4$$

where $\alpha_2 = 6(n-1)(N_2 - n^2) + n^4 + 3N_2^2 - 4nN_3$ and $\alpha_4 = 2n^3 - 2n(n+1) \times N_2 - 6N_2^2 + 8nN_3$. An unweighted test statistic is constructed over families ($f$) as

$$\cdot t_S = \sum_f (\Omega_f - \eta_f)/\sqrt{\sum_f \hat{V}_{Sf}} .$$

## SIMULATION STUDY

The various test statistics were compared by a simulation study. A codominant marker locus with alleles $A, \bar{A}$ in frequencies $p$, $1-p$ was supposed linked to an extent $\lambda$ to a trait locus with alleles $B, \bar{B}$ in frequencies $q$ and $1-q$. Parents were formed with random union of gametes that had frequencies $pq + \Delta$ for $AB$, etc., where $\Delta$ is the linkage disequilibrium between genes $A$ and $B$. For each pair of parents drawn at random, a family of three, four, or five sibs was generated. Sufficient pairs of parents were drawn to provide 100 families, in each of which the $X$ values were not all the same (33 families of size 3, 34 of size 4, and 33 of size 5). The marker locus values were 2, 1, 0 for genotypes $AA, A\bar{A}, \bar{A}\bar{A}$, respectively, while the trait values had a random normal deviate of mean 0 and variance $2q(1-q) \times (1-h)/h$ added to the values 2, 1, 0 for genotypes $BB, B\bar{B}, \bar{B}\bar{B}$, where $h$ is the heritability of the trait. Under this system, for every family, $\mathcal{E}H_3 = \lambda^2[5\Delta^2 + (1-2p)(1-2q)\Delta + 2p(1-p)q(1-q)]$.

For a nominal significance level of 5%, the hypothesis of no linkage was rejected when any test statistic was greater than 1.645. In table 2, we show the proportion of such rejections from 500 replicates of the 100 family situations. In particular,

TABLE 2

EMPIRICAL POWERS ($\Delta = 0, h = 0.5$)

| | | | UNWEIGHTED | | | | WEIGHTED | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $t_{3u}$ | | $t_{6u}$ | | $t_{3w}$ | | $t_{6w}$ | | SMITH |
| $\lambda$ | $p$ | $q$ | N* | E† | N | E | N | E | N | E | $t_S$ |
| 0..... | .5 | .5 | .06 | .06 | .05 | .05 | .05 | .05 | .05 | .06 | .06 |
| | .5 | .9 | .09 | .07 | .06 | .05 | .06 | .05 | .08 | .06 | .04 |
| | .9 | .5 | .05 | .05 | .05 | .05 | .06 | .06 | .06 | .06 | .06 |
| | .9 | .9 | .07 | .07 | .08 | .07 | .09 | .07 | .10 | .08 | .08 |
| .5..... | .5 | .5 | .14 | .14 | .12 | .13 | .16 | .15 | .16 | .16 | .16 |
| | .5 | .9 | .18 | .15 | .15 | .12 | .18 | .14 | .22 | .18 | .17 |
| | .9 | .5 | .16 | .16 | .18 | .18 | .21 | .21 | .22 | .20 | .20 |
| | .9 | .9 | .18 | .15 | .19 | .17 | .22 | .19 | .23 | .21 | .20 |
| .9..... | .5 | .5 | .36 | .36 | .37 | .37 | .45 | .44 | .49 | .48 | .46 |
| | .5 | .9 | .44 | .41 | .41 | .38 | .48 | .41 | .54 | .48 | .43 |
| | .9 | .5 | .45 | .46 | .48 | .48 | .57 | .58 | .57 | .56 | .58 |
| | .9 | .9 | .47 | .43 | .47 | .43 | .56 | .50 | .56 | .52 | .51 |
| 1.0..... | .5 | .5 | .69 | .70 | .73 | .74 | .86 | .86 | .82 | .82 | .86 |
| | .5 | .9 | .76 | .73 | .73 | .70 | .80 | .76 | .82 | .79 | .79 |
| | .9 | .5 | .80 | .79 | .85 | .84 | .93 | .93 | .90 | .90 | .92 |
| | .9 | .9 | .85 | .83 | .88 | .85 | .92 | .90 | .93 | .92 | .92 |

* N: Moments estimated assuming normally distributed trait.
† E: Moments estimated not assuming normally distributed trait.

TABLE 3

EMPIRICAL POWERS ($p = q = 0.5, h = 0.5$)

| | | UNWEIGHTED | | | | WEIGHTED | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $t_{3u}$ | | $t_{6u}$ | | $t_{3w}$ | | $t_{6w}$ | | SMITH |
| $\lambda$ | $\Delta$ | N* | E† | N | E | N | E | N | E | $t_S$ |
| .0..... | −.25 | .09 | .09 | .02 | .04 | .02 | .02 | .09 | .09 | .02 |
| | .00 | .03 | .04 | .05 | .04 | .05 | .04 | .04 | .05 | .03 |
| | +.25 | .00 | .00 | .01 | .01 | .03 | .03 | .06 | .06 | .04 |
| .5..... | −.25 | .58 | .58 | .43 | .43 | .45 | .46 | .68 | .68 | .47 |
| | .00 | .12 | .13 | .10 | .10 | .11 | .12 | .12 | .18 | .12 |
| | −.25 | .57 | .60 | .42 | .41 | .47 | .51 | .65 | .65 | .53 |
| .75 ..... | −.25 | .98 | .98 | .96 | .94 | .98 | .96 | 1.00 | 1.00 | .99 |
| | .00 | .32 | .30 | .32 | .36 | .48 | .47 | .40 | .40 | .48 |
| | +.25 | 1.00 | .99 | .95 | .94 | .96 | .96 | .99 | 1.00 | .98 |
| 1.00 ..... | −.25 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | .00 | .69 | .71 | .71 | .70 | .84 | .86 | .80 | .82 | .85 |
| | +.25 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

* N: Moments estimated assuming normally distributed trait.
† E: Moments estimated not assuming normally distributed trait.

table 2 shows the effects of linkage and gene frequencies when there is no linkage disequilibrium and heritability is .5.

In table 3, we show the effects of linkage disequilibrium in the extreme situation of $\Delta = \pm.25$ when $p = q = .5$. Here the empirical powers are based on 100 replicates. A further set of simulations with the extreme heritability of 1.0 when $p = q = .5$ and $\Delta = 0$, and with 100 replicates, gave the empirical powers for various $\lambda$ values shown in table 4.

## DISCUSSION

The procedure outlined provides a measure within a family of sibs for the detection of linkage between genes at a marker locus and genes affecting a quantitative trait. Families must be of size 3 or larger and must contain at least two distinct genotypes at the marker locus. Thus, all noninformative matings at the

TABLE 4

EMPIRICAL POWERS ($p = q = 0.5, \Delta = 0, h = 1.0$)

| | UNWEIGHTED | | | | WEIGHTED | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $t_{3u}$ | | $t_{6u}$ | | $t_{3w}$ | | $t_{6w}$ | | SMITH |
| $\lambda$ | N* | E† | N | E | N | E | N | E | $t_S$ |
| .0..... | .05 | .05 | .06 | .07 | .10 | .05 | .10 | .07 | .06 |
| .5..... | .58 | .53 | .54 | .51 | .73 | .65 | .76 | .66 | .63 |
| .75..... | .75 | .96 | .95 | .95 | .99 | .97 | 1.00 | .99 | .97 |
| 1.0..... | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

* N: Moments estimated assuming normally distributed trait.
† E: Moments estimated not assuming normally distributed trait.

marker locus are eliminated at the outset. As a consequence of the within-sibship comparison, a positive result cannot stem from fortuitous linkage disequilibria alone. A positive result can stem from direct or pleiotropic effects of the marker gene on the quantitative trait, in which case $\lambda = 1$ and is indistinguishable by this method from tight linkage. It does not matter that genotypes may be in different linkage phases in different families; the quadrivariate nature of the measure dictates a positive contribution with linkage.

Multiple alleles at the marker locus cause no problems as long as there are no more than two alleles within a family. An allele in each family is chosen and scored. With multiple alleles in a family, one can choose the allele with the frequency nearest .5 and ignore the others.

The numerical results point out the low power of these single-generation procedures. The great similarity among the test statistics considered is disappointing. It was thought that weighted ones would be much more powerful, with an additional improvement with the regression approach. There is a very slight improvement with weighting and regression, but $t_{6w}$ is judged to have no advantage of power over Smith's test statistic.

The tests assuming normality or nonnormality also gave very comparable results. The simulated noise variation was normal, but the other variation was not. The computations simplify considerably when normality is assumed, but this is no real advantage when a computer is used, and the estimation of both $\mu_4$ and $\sigma^4$ is recommended, which was what Smith did.

It is somewhat surprising that the assumption that the test statistics behave as normal deviates when $\lambda = 0$ gave roughly the correct rejection percentage, although there is some variation with different test statistics and parameters.

The results for different gene frequencies may at first appear surprising. One would expect the powers to decrease as gene frequencies deviate from .5. The expected frequency of pairs of homozygous parents for the $X$'s is $[p^2 + (1 - p)^2]^2 = .25$ for $p = .5$ and .67 for $p = .9$. Consequently, many more families had to be generated for $p = .9$ than for $p = .5$ to have 100 with variation in the $X$'s. Had the number of families generated been held constant, the tests would have been based on many fewer families for $p = .9$. The fact that the power did not decrease for $q = .9$ as opposed to $q = .5$ is accounted for with a different reason. The noise variable added into $Y$ had a variance of $2q(1 - q)(1 - h)/h$. With $h = .5$, this variance is .5 for $q = .5$ and .18 for $q = .9$. The reduction in informative families for $Y$ from increasing $q$ to .9 is offset by reduction in the noise variance. The noise variance and gene frequencies could be varied independently of each other for a fixed number of families generated, but we believe we have given a sufficiently general picture of the results.

Power, of course, increases with heritability and linkage. Linkage disequilibrium enhances the power but does not mimic the effects of linkage since within-family statistics are utilized.

On balance, we are forced to conclude that the use of sibs for detecting linkage between a marker locus and a quantitative trait is unlikely to be of great practical value. When information is available only on families of sibs, the procedure may be used as an indicator of linkage.

## APPENDIX A

## VARIANCES OF $H_3$ AND $H_4$

For a family, let

$$z' = [\sum_i x_i^2 y_i^2 , \quad \sum_i x_i^2 \sum_i y_i^2 , \quad (\sum_i x_i y_i)^2] ,$$

so that $H_3 = w_3'z$, $H_4 = w_4'z$, with

$$w_3' = \left[ \frac{1}{n-1} , \quad \frac{-(n+2)}{n(n-1)(n-2)} , \quad \frac{4}{n(n-3)} \right]$$

$$w_4' = \left[ \frac{-4}{(n-2)(n-3)} , \quad \frac{4}{n(n-1)(n-2)(n-3)} , \quad \frac{4}{n(n-3)} \right] .$$

Dropping the 3 or 4 subscripts, since the same argument holds in each case, $\text{var}(H) = w' \mathscr{E} z z' w$. Required in the evaluation of $\mathscr{E} z z'$ are the expectations of fourth-order functions of the $y$'s in terms of $\mu_4$ and $\sigma^4$. The coefficients of the moments in these expectations follow.

| | $\mu_4$ | $\sigma^4$ |
|---|---|---|
| $\mathscr{E} y_i^4$ ..... | $\frac{(n-1)(n^2-3n+3)}{n^3}$ | $\frac{3(n-1)(2n-3)}{n^3}$ |
| $\mathscr{E} y_i^3 y_j$..... | $\frac{-(n^2-3n+3)}{n^3}$ | $\frac{-3(2n-3)}{n^3}$ |
| $\mathscr{E} y_i^2 y_j^2$ ..... | $\frac{2n-3}{n^3}$ | $\frac{n^3-2n^2-3n+9}{n^3}$ |
| $\mathscr{E} y_i^2 y_j y_k$ ..... | $\frac{n-3}{n^3}$ | $\frac{-(n+3)(n-3)}{n^3}$ |
| $\mathscr{E} y_i y_j y_k y_l$..... | $\frac{-3}{n^3}$ | $\frac{3(n+3)}{n^3}$ |

The evaluation is now straightforward but very tedious.

$$\mathscr{E} zz' = P \sum_i x_i^4 + Q(\sum_i x_i^2)^2 ,$$

where

$$P = \begin{bmatrix} \frac{(n-2)A}{n^2} & 0 & \frac{A}{n} \\ 0 & 0 & 0 \\ \frac{A}{n} & 0 & \mu_4 - 3\sigma^4 \end{bmatrix}$$

and

$$Q = \begin{bmatrix} \frac{A+B}{n^3} + \frac{(n-3)\sigma^4}{n} & \frac{(n-1)B}{n^2} & \frac{B-A}{n^2} \\ \frac{(n-1)B}{n^2} & \frac{(n-1)B}{n} & \frac{B}{n} \\ \frac{B-A}{n^2} & \frac{B}{n} & 3\sigma^4 \end{bmatrix}$$

with $A = (n-2)\mu_4 - (n-6)\sigma^4$, $B = (n-1)\mu_4 + (n^2-2n+3)\sigma^4$.

## APPENDIX B

## ESTIMATION OF THE MOMENTS $\mu_4$ AND $\sigma^4$

The method of estimation is motivated by the following two expectations: $\mathscr{E}(y_i - y_j)^4 = 2(\mu_4 + 3\sigma^4)$ and $\mathscr{E}(y_i - y_j)^2(y_k - y_l)^2 = 4\sigma^4$. Then, the sums

$$M_1 = \sum_{i \neq j} \sum (y_i - y_j)^4 = 6(\sum_i y_i^2)^2 + 2n \sum_i y_i^4$$

$$M_2 = \sum_{i \neq j} \sum_{k \neq l} \sum (y_i - y_j)^2 (y_k - y_l)^2$$

$$= 4(n^2 - 3n + 3)(\sum_i y_i^2)^2 - 4n(n-1) \sum_i y_i^4$$

have expectations $\mathscr{E} M_1 = 2n(n-1)(\mu_4 + 3\sigma^4)$ and $\mathscr{E} M_2 = 4n(n-1)(n-2)(n-3)\sigma^4$. Estimates are obtained from combining information over families ($f$)

$$\hat{\sigma}^4 = \sum_f [(n_f^2 - 3n_f + 3)(\sum_i y_{fi}^2)^2 + n_f(n_f - 1) \sum_i y_{fi}^4] / \sum_f n_f$$

$$\times (n_f - 1)(n_f - 2)(n_f - 3)$$

$$\hat{\mu}^4 = \sum_f [3(\sum_i y_{fi}^2)^2 + n_f \sum_i y_{fi}^4] / \sum_f n_f(n_f - 1) - 3\hat{\sigma}^4 .$$

When there is reason to believe the trait is normally distributed, only one estimator is needed:

$$\hat{\sigma}^2 = \sum_f \sum_i y_{fi}^2 / \sum_f (n_f - 1) ,$$

with $\hat{\sigma}^4 = (\hat{\sigma}^2)^2$, $\hat{\mu}_4 = 3(\hat{\sigma}^2)^2$.

Either set of estimators is substituted into the expressions of APPENDIX A to provide estimates of the variances of $H_3$ and $H_4$.

## ACKNOWLEDGMENT

We thank Mr. John Hackman for computing assistance.

## REFERENCES

1. COCKERHAM CC: Sib pairing methodology, in *Genetic Analysis of Common Diseases: Applications to Predictive Factors in Coronary Disease*, edited by SING CF, SKOLNICK M, New York, Alan R. Liss, 1979, pp 417–430
2. PENROSE LS: Genetic linkage in graded human characters. *Ann Eugen* 88:233–237, 1937
3. SMITH CAB: A non-parametric test for linkage with a quantitative character. *Ann Hum Genet* 38:451–460, 1975
4. BENER A, LOESCH D, SMITH CAB: Simplified formulas for detecting linkage with a quantitative character. *Ann Hum Genet* 43:249–254, 1980