

THE INFLUENCE OF PARTIAL MANIFESTATION ON THE DETECTION OF LINKAGE

NORMAN T. J. BAILEY

Department of Medicine, University of Cambridge

Received 20.vii.49

I. INTRODUCTION

IN the usual type of mating, involving two or more factors, where we can estimate the intensity of linkage from the proportion of recombinant types, there are a number of disturbing influences which prevent a simple interpretation of the data. The most well known are probably the occurrence of differential viability and the possibility of differential recombination fractions in male and female heterozygotes. The use of linkage backcrosses in both coupling and repulsion to overcome these difficulties is now a standard procedure. Another source of disturbance of the observed numbers is partial manifestation. Organisms, heterozygous for a dominant gene or homozygous for a recessive gene, may fail to exhibit the character in question in a proportion of cases and thus be effectively misclassified. Fairfield Smith (1937) considered the problem of estimating the degree of misclassification, but of much greater importance is the question how much are the usual *estimates of linkage* affected. We shall assume, for the most part, that differential viability is absent, as disturbances due to this cause are often small and require a large number of animals to reveal their presence, whereas partial manifestation may be quite considerable in fairly small bodies of data. There are two main methods of coping with this sort of complication: one can either try to set up matings showing full manifestation; or one can develop the statistical analysis to deal with the difficulty. With experimental work at any rate one can employ both methods, although with human genetics only the latter course is open to us.

The basic problem is this: we have a character in which we are interested which appears to have imperfect penetrance—we have noticed skipped generations or a characteristic disturbance of the observed numbers—and we wish to examine the possibility of linkage to other genes, which may *themselves* not have complete manifestation. If linkage proves to be significant then we shall want to estimate it. A point of particular importance in the latter connection is the use of a method which will give an efficient estimate of the linkage when the degree of manifestation varies from one group of matings to another.

We shall be concerned mainly with methods applicable to the usual type of experimental data ; and shall conclude with a few illustrations of the treatment of population data such as are collected in the study of human genetics.

2. NOTE ON SCORING WITH SEVERAL PARAMETERS

Suppose the expectations involve several parameters : $p ; \theta_1, \theta_2 \dots$ and we are particularly concerned with the significance and estimation of one of them, say p . Then if L is the logarithm of the likelihood, an efficient score for $p = p_1$ is $S(p_1) = \frac{\partial L}{\partial p} \Big|_{p = p_1}$, where the maximum likelihood estimates of the parameters : $\theta_1, \theta_2 \dots$, given $p = p_1$, have been substituted. It is easy to show that the amount of "partial" information about p is the reciprocal of the corresponding element in the inverse of the complete information matrix.

For large samples the appropriate test of significance is the χ^2 with one degree of freedom :

$$\chi^2 = \frac{S^2}{I_p} \dots \dots \dots (2.1)$$

where I_p is the partial information about p .

On the other hand we can find the maximum likelihood estimate of p by the usual method of interpolating between two sufficiently close scores of opposite sign. In this case the easiest way of calculating I_p is to use the formula

$$I_p \doteq \frac{S(p_1) - S(p_2)}{p_2 - p_1} \dots \dots \dots (2.2)$$

where the score has been calculated at adjacent values, p_1 and p_2 .

We have of course :—

$$\text{Var } p = I_p^{-1} \dots \dots \dots (2.3)$$

These results have been used to derive the tests of significance and the methods of estimation employed below.

3. LINKAGE BACKCROSS IN TWO FACTORS :

$$AB/ab = ab/ab$$

Let us consider a linkage backcross with two factors in coupling of the type : $AB/ab = ab/ab$ where A is misclassified as a in a proportion λ of cases. We can call the latter : $A \rightarrow a$. We can use the formulæ given below for matings in repulsion and for misclassifications of the type, $a \rightarrow A$, by a suitable transposition of the symbols. Let the observed numbers in the four phenotypic classes be :—

AB	Ab	aB	ab	Total
a	b	c	d	n

The usual test for linkage in the absence of disturbing influences is :—

$$\chi^2 = \frac{\{(b+c)-(a+d)\}^2}{n} \quad . \quad . \quad . \quad (3.1)$$

If, however, we admit partial manifestation χ^2 is found to assume the modified form :—

$$\chi^2 = \frac{4(bc-ad)^2}{n(a+b)(c+d)} \quad . \quad . \quad . \quad (3.2)$$

We can clearly combine χ^2 obtained from families with different degrees of manifestation.

The maximum likelihood estimates for λ and p , the recombination fraction, are given by :—

$$\lambda = \frac{(cd-ab)}{(a+c)(b+d)} \quad . \quad . \quad . \quad (3.3)$$

$$p = \frac{b(a+c)}{b(a+c)+a(b+d)} \quad . \quad . \quad . \quad (3.4)$$

The variances of these estimates are :—

$$\text{var } \lambda = \frac{2(1-\lambda)}{n} \{2pq + \lambda(1-2pq)\} \quad . \quad . \quad . \quad (3.5)$$

$$\text{var } p = \frac{2pq}{n} \left(\frac{1}{1-\lambda} - 2pq \right) = \frac{2pq}{n} (\theta - 2pq) \quad . \quad . \quad (3.6)$$

$$\text{where } \theta = \frac{1}{1-\lambda}.$$

These formulæ assume that we can take λ as the same throughout the whole body of data. We shall consider below the problem of combining estimates from groups with different λ 's.

Linkage matings will as a rule give more information about the existence of partial manifestation than those without linkage. In the present case the test for manifestation, admitting linkage, is :—

$$\chi^2 = \frac{4(ab-cd)^2}{n(a+d)(b+c)} \quad . \quad . \quad . \quad (3.7)$$

The amount of information about λ for small λ is n in a single backcross, but $\frac{n}{4pq}$ in a linkage backcross. There is a gain since $4pq < 1$ unless $p = q = \frac{1}{2}$.

For other types of mating, *i.e.* repulsion and $a \rightarrow A$, we have the following changes. In each case the four classes are labelled (a), (b), (c) and (d) in turn.

TABLE I

	(a)	(b)	(c)	(d)	
(1) AB/ab = ab/ab ; A → a	AB	Ab	aB	ab	p is estimated in each case.
(2) Ab/aB ; A → a	Ab	AB	ab	aB	
(3) AB/ab ; a → A	ab	aB	Ab	AB	
(4) Ab/aB ; a → A	aB	ab	AB	Ab	

Example

The data used here is taken from Professor Fisher's material involving the third chromosome of the house mouse. Pirouette, pi , is a simple recessive; and light head, te , is a recessive in which the homozygote, te/te , does not always show the character. The matings are in coupling, *i.e.* $+/te\ pi = te\ pi/te\ pi$. This is clearly type (3) above. The observed numbers are:—

$te\ pi$	$te\ +$	$+ \pi i$	$++$	Total
40	16	93	102	251
a	b	c	d	n

From (3.2) we find $\chi^2 = 9.82$, while (3.4) and (3.6) give $p = 31.1 \pm 6.6$ per cent.

4. ESTIMATE FROM THE MANIFESTING CLASSES ONLY

Before passing on to a more complicated case let us consider an approximation for p .

It is obvious that we can obtain a consistent estimate of the recombination fraction by considering the manifesting classes alone, *i.e.*

$$\check{p} = \frac{b}{a+b} \quad . \quad . \quad . \quad . \quad (4.1)$$

The expected number of individuals in the manifesting classes is $\frac{1}{2}n(1-\lambda)$ and so the variance of the above estimate is:—

$$\text{var } \check{p} = \frac{2pq}{n(1-\lambda)} \quad . \quad . \quad . \quad . \quad (4.2)$$

Comparing this with the variance of the estimate obtained when we employ all four classes, we can define the efficiency, E , as the ratio of the two quantities.

$$\text{Therefore } E = 1 - 2pq(1-\lambda) \quad . \quad . \quad . \quad (4.3)$$

It is easy to see that the efficiency increases with close linkage and low penetrance. For 10 per cent. recombination, $E = 0.82$ for complete penetrance and is greater than 0.91 for $\lambda > 0.5$.

5. COMBINING ESTIMATES OF p FROM DATA WITH VARYING DEGREES OF MANIFESTATION

This is a situation frequently met with in practice. Suppose the data falls naturally into groups with different degrees of manifestation. It is convenient to use:—

$$\theta = \frac{1}{1-\lambda}; \quad 1 \leq \theta \leq \infty \text{ for } 0 \leq \lambda \leq 1 \quad . \quad . \quad (5.1)$$

Applying the results given in section (2) above we find that for a given group the score for p is:—

$$S_p = -\frac{a}{q} + \frac{b}{p} + \frac{c}{\theta-q} - \frac{d}{\theta-p} \quad . \quad . \quad . \quad (5.2)$$

where θ is given by the quadratic :—

$$(a+b)\theta^2 - (a+b+qc+pd)\theta + npq = 0 \quad . \quad . \quad (5.3)$$

The amount of partial information about p is readily obtained from adjacent values of the score.

Example

Let us take another example from the third chromosome of the house mouse. This time we will consider backcrosses in coupling for luxate, lx , and dominant spotted, W . Luxate is here run as a dominant and often fails to manifest. Thus the matings are :—

$$lx\ W/++ = ++/++$$

i.e. type (1).

The observed numbers and the appropriate calculations are set out below :—

	(a)	(b)	(c)	(d)	Total
	lxW	$lx+$	$+W$	$++$	
1.	24	7	65	69	165
2.	37	9	38	66	150
3.	19	2	12	30	63

p	Quadratic for θ (≥ 1)	θ	Score	I
1. 0.18	$31\theta^2 - 96.72\theta + 24.3540 = 0$	2.84374	+ 15.8359	259.39
0.19	$31\theta^2 - 96.76\theta + 25.3935 = 0$	2.83205	+ 13.2420	
2. 0.18	$46\theta^2 - 89.04\theta + 22.1400 = 0$	1.64265	+ 5.9467	386.54
0.19	$46\theta^2 - 89.32\theta + 23.0850 = 0$	1.63475	+ 2.0813	
3. 0.18	$21\theta^2 - 36.24\theta + 9.2988 = 0$	1.41214	- 16.1420	132.37
0.19	$21\theta^2 - 36.42\theta + 9.6957 = 0$	1.40588	- 17.4657	

Summary

$p = 0.18$	Score	$p = 0.19$	I_p	$1-\lambda$
1. + 15.8359		+ 13.2420	259.39	0.35
2. + 5.9467		+ 2.0813	386.54	0.61
3. - 16.1420		- 17.4657	132.37	0.71
<hr/>		<hr/>	<hr/>	
+ 5.6406		- 2.1424	778.30	

Interpolating gives $p = 0.1872$ and $(778.30)^{-1} = 0.0358$.

Therefore $p = 18.72 \pm 3.58$ per cent.

The estimate from the manifesting classes alone is 18/98, *i.e.* $p = 18.37$ per cent. The amount of information is now only 653.6, giving a standard error of 3.91 per cent.—this estimate has an efficiency of 84 per cent.

6. LINKAGE TO A PARTIALLY MANIFESTING CHARACTER :

$$AB/ab = ab/ab ; A \rightarrow a, B \rightarrow b$$

Suppose that B, instead of being a normal dominant character, also has incomplete manifestation of the type : $B \rightarrow b$, and let the proportion misclassified be ν . Then the test for linkage is :—

$$\chi^2 = \frac{n(bc-ad)^2}{(a+b)(a+c)(b+d)(c+d)} \quad . \quad . \quad . \quad (6.1)$$

This is immediately clear from the remark that if there is no linkage then the expected numbers are in the following proportions :—

$$(1-\lambda)(1-\nu) : (1-\lambda)(1+\nu) : (1+\lambda)(1-\nu) : (1+\lambda)(1+\nu)$$

and these can be rearranged in the form of a 2×2 contingency table.

The estimates of p , λ and ν are :—

$$p = 1 - \frac{an}{2(a+b)(a+c)} ; \lambda = \frac{(c+d)-(a+b)}{n} ; \nu = \frac{(b+d)-(a+c)}{n} \quad (6.2)$$

We also have :—

$$\text{var } p = an \frac{\{n(a^2+bc) - a(a+b)(a+c)\}}{4(a+b)^3(a+c)^3} \quad . \quad . \quad (6.3)$$

The various types of mating, with the appropriate modifications, are :—

TABLE 2

	(a)	(b)	(c)	(d)	Estimate
(1) $AB/ab = ab/ab ; A \rightarrow a, B \rightarrow b$	AB	Ab	aB	ab	p
(2) $Ab/aB \quad A \rightarrow a, B \rightarrow b$	AB	Ab	aB	ab	q
(3) $AB/ab \quad a \rightarrow A, B \rightarrow b$	aB	ab	AB	Ab	q
(4) $Ab/aB \quad a \rightarrow A, B \rightarrow b$	aB	ab	AB	Ab	p
(5) $AB/ab \quad a \rightarrow A, b \rightarrow B$	ab	aB	Ab	AB	p
(6) $Ab/aB \quad a \rightarrow A, b \rightarrow B$	ab	aB	Ab	AB	q

With types (2), (3) and (6), (6.2) gives an estimate of q instead of p . The formulæ for λ and ν are unaffected.

Example

Let us make a direct estimate of the linkage between luxate and light head. The matings are : $+lx/te+ = te+/te+$; $te \rightarrow +, lx \rightarrow +$, *i.e.* type (3). The observed numbers are :—

$te \ lx$	$te \ +$	$+ \ lx$	$++$	Total
7	49	41	154	251
a	b	c	d	n

Applying the above results we obtain $\chi^2 = 2.1$. This is not significant, but as the genes are known to be linked we can proceed to estimate the recombination fraction and, using (6.2) and (6.3), find : $p = 32.7 \pm 10.6$ per cent.

7. DIFFERENTIAL VIABILITY

Suppose we return to the case of A alone having partial manifestation, and wish to discover whether part of the discrepancy in the observed ratios is due to differential viability. Without a linked locus it is not possible to distinguish these two effects statistically. But if we can find a pair of linked genes, B and b, unaffected by partial manifestation or differential viability, then we are in a position to make the test, for the three degrees of freedom available between the four observed classes can now be allotted to linkage, manifestation and viability. We find :—

$$\chi^2 = \frac{\{(a+c)-(b+d)\}^2}{n} \quad \dots \quad (7.1)$$

which curiously enough is the χ^2 we should obtain if we were testing the segregation of B/b alone !

If differential viability is present then we can estimate p from the manifesting classes alone.

8. HUMAN FAMILIES

Now let us consider one or two examples of the treatment of data from human families. We shall use certain modifications of the method of u -statistics developed by Fisher (1935a, 1935b and 1936), and extended by Finney (1940) to cover a wide variety of instances. We will assume for simplicity that we have a test factor with dominance, represented by T/t, and that the frequencies of the allelomorphous genes for the test factor are unknown. In this case families with no tt children must be rejected. The analysis can be extended to the case where the frequencies are known in the manner described by Finney.

(a) “ Dominant ” abnormalities

Let us take the case of a so-called dominant abnormality, like diabetes insipidus, which sometimes skips generations and is assumed to be due to a single gene, D, with incomplete manifestation. The observed numbers are :—

DT	Dt	dT	dt	Total
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>n</i>

The use of Fisher’s u -statistics gives the following familiar results for the scores and amounts of information with respect to a quantity $\zeta = 1 - 4\chi(1 - \chi)$, where χ is the recombination fraction :—

(i) Double backcross, DT = dt.

$$\begin{aligned} u_{11} &= (a - b - c + d)^2 - (a + b + c + d) \\ S &= \Sigma \frac{1}{2} u_{11} ; I = \Sigma \frac{1}{2} n(n - 1) \end{aligned} \quad \dots \quad (8.1)$$

(ii) Single backcross (intercross for test factor), $DT = dT$.

$$\left. \begin{aligned} u_{31} &= (a-3b-c+3d)^2 - (a+9b+c+9d) \\ S &= \Sigma \frac{1}{18} u_{31}; I = \Sigma \frac{1}{18} n(n-1) \frac{4^n - 3^{n-2}}{4^n - 3^n} \end{aligned} \right\} \dots (8.2)$$

where Σ indicates summation over families.

Now suppose that a proportion λ of the abnormal are misclassified as normal. The appropriate modifications of the above results are, writing $\mu = \frac{1-\lambda}{1+\lambda}$:—

$$\left. \begin{aligned} \text{(i) } DT &= dt : D \rightarrow d. \\ u'_{11} &= (a-b-\mu c + \mu d)^2 - (a+b + \mu^2 c + \mu^2 d) \\ &= \{(a-b)^2 - (a+b)\} - 2\mu(a-b)(c-d) \\ &\quad + \mu^2\{(c-d)^2 - (c+d)\} \\ S' &= \Sigma \frac{1}{2} \mu'_{11}; I' = \mu^2 I = \Sigma \frac{1}{2} \mu^2 n(n-1) \end{aligned} \right\} \dots (8.3)$$

$$\left. \begin{aligned} \text{(ii) } DT &= dT; D \rightarrow d. \\ u'_{31} &= (a-3b-\mu c + 3\mu d)^2 - (a+9b + \mu^2 c + 9\mu^2 d) \\ &= \{(a-3b)^2 - (a+9b)\} - 2\mu(a-3b)(c-3d) \\ &\quad + \mu^2\{(c-3d)^2 - (c+9d)\} \\ S' &= \Sigma \frac{1}{18} u'_{31}; I' = \mu^2 I = \Sigma \frac{1}{18} \mu^2 n(n-1) \frac{4^n - 3^{n-2}}{4^n - 3^n} \end{aligned} \right\} \dots (8.4)$$

In each case the new score is a simple modification of the old and the amount of information is multiplied by a factor μ^2 . μ is estimated by $\frac{\Sigma(a+b)}{\Sigma(c+d)}$ in both (i) and (ii). Thus in practice when we have data from both kinds of families we can estimate μ from the totals. If there is reason to suspect that μ differs in different sections of the data then the scores and amounts of information can be calculated for the various groups on the basis of the appropriate value of μ before combining. In the present case the scores are independent of the estimates of manifestation so we need consider only the u -statistics. It is worth noting that we shall be able to subsume under the present head families in which one parent, although not showing the character, is known from the pedigree to be heterozygous, provided of course that ascertainment is virtually through the parents. If ascertainment is through the children a slightly different treatment is required.

(b) Recessive abnormalities

Let us now consider a recessive defect in which the homozygous recessive sometimes fails to manifest, like the Laurence-Moon-Biedl syndrome. We will take two illustrations, both intercrosses with respect to the defect, one an intercross and one a backcross for the test factor. These cases are a little more complicated in that the

score for linkage is not now independent of manifestation and we are obliged to calculate the three information functions $I_{\zeta\zeta}$, $I_{\zeta\lambda}$ and $I_{\lambda\lambda}$. Ascertainment is assumed to be through the children. The observed numbers are :—

TR	Tr	tR	tr	Total
a	b	c	d	n

The modified results are as follows, where now $\mu = \frac{1-\lambda}{3+\lambda}$:—

$$\left. \begin{aligned}
 \text{(i) TR} &= tR ; r \rightarrow R. \\
 \frac{1}{9}u'_{31} &= (\mu a - b - \mu c + d)^2 - (\mu^2 a + b + \mu^2 c + d) \\
 &= \mu^2 \{ (a-c)^2 - (a+c) \} - 2\mu(a-c)(b-d) \\
 &\quad + \{ (b-d)^2 - (b+d) \} \\
 \text{where } \mu &= \frac{\Sigma(b+d-1)}{\Sigma(a+c)}. \\
 S &= \Sigma \frac{1}{8} u'_{31}. \\
 I_{\zeta\zeta} &= \Sigma \frac{1}{2} (n-1) \{ (n-2)\mu^2 + 2\mu \} \\
 I_{\lambda\lambda} &= \Sigma \frac{(n-1)}{(1-\lambda)(3+\lambda)} \\
 I_{\zeta\lambda} &= \Sigma \frac{(n-1)}{(3+\lambda)(2^n-1)}
 \end{aligned} \right\} \dots (8.5)$$

$$\left. \begin{aligned}
 \text{(ii) TR} &= TR ; r \rightarrow R. \\
 \frac{1}{9}\mu'_{33} &= (\mu a - b - 3\mu c + 3d)^2 - (\mu^2 a + b + 9\mu^2 c + 9d) \\
 &= \mu^2 \{ (a-3c)^2 - (a+9c) \} - 2\mu(a-3c)(b-3d) \\
 &\quad + \{ (b-3d)^2 - (b+9d) \} \\
 \text{where } \mu &= \frac{\Sigma(b+d-1)}{\Sigma(a+c)}. \\
 S &= \Sigma \frac{1}{8} \mu'_{33}. \\
 I_{\zeta\zeta} &= \Sigma \frac{2}{9} (n-1) \{ (n-2)\mu^2 + 2\mu \} \frac{4^n - 3^{n-2}}{4^n - 3^n} \\
 I_{\lambda\lambda} &= \Sigma \frac{(n-1)}{(1-\lambda)(3+\lambda)} \\
 I_{\zeta\lambda} &= \Sigma \frac{2 \cdot 3^{n-2} (n-1)}{(3+\lambda)(4^n - 3^n)}
 \end{aligned} \right\} \dots (8.6)$$

The test for linkage, given partial manifestation, is :—

$$\left. \begin{aligned}
 \chi^2 &= S^2 / I_{\zeta} \\
 \text{where } I_{\zeta} &= I_{\zeta\zeta} - \frac{I_{\zeta\lambda}^2}{I_{\lambda\lambda}}
 \end{aligned} \right\} \dots (8.7)$$

There are, of course, a very large number of different cases which might arise in practice and the above examples are merely intended to illustrate the general method of approach.

I am indebted to Professor R. A. Fisher for many useful discussions and for providing me with the data on the 3rd chromosome of the house mouse employed in the examples.

9. REFERENCES

- FINNEY, D. J. 1940. The detection of linkage. *Ann. Eug.*, 10, 171-214.
- FISHER, R. A. 1935a. The detection of linkage with "Dominant" abnormalities. *Ann. Eug.*, 6, 187-201.
- FISHER, R. A. 1935b. The detection of linkage with recessive abnormalities. *Ann. Eug.*, 6, 339-351.
- FISHER, R. A. 1936. Tests of significance applied to Haldane's data on partial sex-linkage. *Ann. Eug.*, 7, 87-104.
- SMITH, H. FAIRFIELD. 1937. The test of significance for Mendelian ratios when classification is uncertain. *Ann. Eug.*, 8, 94-95.