THE OHIO JOURNAL OF SCIENCE

VOL. XXXIX NOVEMBER, 1939 No. 6

SOME PRACTICAL CONSIDERATIONS IN TESTING FOR GENETIC LINKAGE IN SIB DATA

L. S. PENROSE
Colchester, England

The detection of an association between two inherited characters in family data has sometimes erroneously been interpreted as evidence of genetic coupling or linkage. To establish linkage an equivalent degree of dissociation between the two characters must be demonstrated in the families where the genes which underlie the characters are in repulsion. It has, for example, been asserted that the coincidence of retinitis pigmentosa and polydactyly in certain sibships may be due to the coupling of two genes, one of which underlies the retinal changes and the other the polydactyly (Cockayne, 1935). This hypothesis will be rendered plausible only when sibships are discovered in which polydactyly occurs in just those sibs who are free from retinitis pigmentosa. Otherwise the most likely explanation is that a single gene is responsible for both abnormalities. If an observed association between two characters is imperfect, as is usually the case, this incomplete concomitance need not be interpreted in terms of crossing over. Variation in the expression of the same gene in different members of a population and even among members of the same sibship is common in human data.

It is of some interest to examine the consequences of different types of association of genetic characters. In the first place, if two characters are due to the same gene, positive correlation is observed whether subjects are chosen within specified sibships or selected at random from the general population. The correlation, moreover, is independent of whether or not a physiological connection between the peculiarities can be traced. This type of association must be distinguished from that due to independent genes which have a racial or local grouping. In Holland (Van Herwerden, 1930), for example, a positive correlation, likely to be due to racial grouping, was noted between the possession of agglutinogen B and dark hair color in the general population. In such instances, although the traits are associated in the general population, there will be no correlation between them within the sibships.

Two characters will be found to be negatively correlated within sibships and also in the general population if they are due to allelic genes. This is because the presence of one allelic gene excludes the presence of any other allele. For example, the fact that blood group AB is found to be much less frequent than it should be on the assumption that A and B are independently assorted is, by itself, strong evidence that A and B are allelic genes. Finally it is possible for the same

TABLE I

Relationship Between Two Characters	Correlation Between the Two Characters	
	In the General Population	Within the Sibship
Independence	0	0
Due to same gene	+	+
Racially grouped	+	0
Due to allelic genes	—	—
Physiological compensation	—	—
Genetic linkage	0	+ or —

gene to have effects of a physiologically compensatory nature. Thus, for instance, head shape may vary according to genetic constitution in persons whose head sizes are of the same magnitude: length and breadth of the head may thus be found to be negatively correlated both in sibships and in the general population.

When dealing with the metrical characters which have multiple genetic backgrounds, the possibility of finding conditions which closely simulate linkage in sibships must not be overlooked. If the two characters are totally uncorrelated in sibships the danger of making a mistake is slight. Even so, it is theoretically possible for concomitant and compensatory variations which are not in any way due to genetic linkage to coexist. In Table I the types of association and dissociation which may be mistaken for linkage are set out.

The peculiar feature of genetically linked characters is that they are equally likely to be positively or negatively correlated within single sibships.

The simplest way of testing for genetic linkage between dominant and recessive characters (Penrose, 1935) is to classify all the pairs of sibs in the available data. A 2 × 2 table is compiled in which each pair is entered according to whether its members are like or unlike with respect to the two test characters. A significant deviation from proportionality in such a table can be caused by linkage between the two characters. A similar deviation can also be caused by any of the situations discussed in the paper which give rise to a positive or negative association. These possibilities must therefore always be examined before claiming a positive result in any linkage study.

TABLE II

		CHARACTER A		
		Like	Unlike	Total
	Like	x	y	x + y
CHARACTER B	Unlike	z	w	z + w
	Total	x + z	y + w	t

If the numbers in one or more of the cells in the 2 × 2 table are small, the exact test for significance of a deviation from proportionality can conveniently be applied (Fisher, 1934). An estimate of the recombination value of the two genes can be obtained from the 2 × 2 table by calculating the function ϕ, from Table II, by the following formula

$$\phi = \frac{xw - yz}{(y + w)\ (z + w)} \cdot$$

If there is no crossing over, ϕ has the maximum value of $\frac{1}{2}$ with common characters and of $\frac{1}{3}$ if one or both characters are rare. When there is random assortment, $\phi = 0$. The standard error of ϕ is approximately $\sqrt{1/w - 1/t}$ and so the statistical value of the result depends mainly upon the number of pairs of sibs unlike for both the characters. Unless the investigator is fortunate enough to obtain a small sample of data in which ϕ happens to have an unusually large value, no significant positive result is to be expected unless more than 100 pairs are classified. If sibships in which one of the characters does not occur are left out of the reckoning, there is no loss of information because these sibships cannot contain any pairs unlike in both characters. In practice it is convenient to select sibships by the presence of a comparatively rare dominant or recessive trait in at least one member; tests for linkage are then made with other traits which are found distributed in these sibships. If the second test factor has a fairly high gene frequency in the general population, the cells x, y, z and w in the 2 $\times$ 2 table will all be of the same order of magnitude. In this case significance of a deviation from proportionality can be safely examined by the symmetrical function χ^2, by the formula

$$\chi^2 = \frac{(xw - yz)^2 \, t}{(x + y)\,(x + z)\,(z + w)\,(y + w)} \cdot$$

If the numbers in the four cells are exactly equal in magnitude this test is equivalent to applying Fisher's u_{11} function to the data.

The method of classifying like and unlike pairs of sibs needs amplification if the data concern characters which have more than two alternative values. Data analyzed by Burks (1937) concerning tooth deficiency and hair color are of this type. Metrical characters can be dealt with by differences in sib pairs (Penrose, 1938); they can also be simplified by choosing a point in the scale so that every individual can be classed as greater or less than a given value. It is not permissible to classify sibs so that unlikenesses of more than one kind are grouped together. With multiple alleles care is also needed. Suppose, for example, that in a certain family there are four sibs, as follows: M, A; N, A; MN, a; and MN, a. Here M and N. represent blood agglutinogens and A is a rare recessive trait. The sibships can be scored for like and unlike pairs for the two characters N and A or for M and A separately. The result is the same in either case (Table III) and correctly indicates free assortment of the characters.

If an attempt is made to group like and unlike pairs without specifying two alternatives and only two, a different distribution is obtained (Table IV). There is now a positive deviation from proportionality. The correct way to compile this table is to

TABLE III

		A	
		Like	Unlike
M	Like	1	2
	Unlike	1	2

TABLE IV

		A	
		Like	Unlike
M, N, or MN	Like	1	0
	Unlike	1	4

TABLE V

		A	
		Like	Unlike
MN	Like	1	0
	One degree of unlikeness	0	4
	Two degrees of unlikeness	1	0

distinguish between unlikeness of M and N (two degrees) and unlikeness of MN and M (one degree). The table should be a 2×3 table (Table V) in which symmetry is again established.

The method of counting like and unlike pairs for characters which can take several alternative values will give rise to misleadingly positive results if uncritically applied.

The search for interrelationship due to genetic linkage between psychological factors has been suggested by Price (1937). The same care is needed in selecting suitable mental factors for testing as in choosing physical characters. Only abilities which are uncorrelated can be rightly chosen for the test. If the simple method of like and unlike pairs is to be used, each sib must be scored for presence or absence of each given quality.

REFERENCES

Burks, B. 1937. Carnegie Inst. Wash. Yearbook, 312–319.

Cockayne, E. A., Krestin, D., and **Sorsby, A.** 1935. Quart, Jour. Med., new series 4: 93.

Fisher, R. A. 1934. Statistical Methods, 5th ed. Edinburgh, Oliver and Boyd.

Penrose, L. S. 1935. Annals of Eugenics 4: 133; 1938, ibid, 8: 233.

Price, B. 1937. Psych. Rev. 44: 183.

Van Herwerden, M. A., and **Boele-Nyland, Th. Y.** 1930. Proc. Kon. Akad. Wet. 33: 659.