Ann. Hum. Genet., Lond. (1971), **35**, 47
Printed in Great Britain

Frequency in relatives for an all-or-none trait

BY J. W. JAMES*
Institute of Animal Genetics, Edinburgh

The frequency of an all-or-none trait in the relatives of individuals possessing the attribute may be calculated for a simple genetic model by the direct method of constructing frequency tables for all genotypes in index cases and their relatives. This is often tedious, and the systematic matrix method of Li & Sacks (1954) may be more convenient, as shown by Elston & Campbell (1970). Our purpose here is to present another approach and to consider some of its implications.

Suppose individuals are taken at random from the whole population and given a score of 1 if they have the trait and 0 if they do not. Let the variable X denote the scores of these primary individuals. Suppose that the relatives of the primary individuals are scored in the same way, the variable Y denoting the scores on relatives. Then if K_P is the proportion of the population which has the attribute, X and Y have the same mean, K_P, and they also have the same variance, $K_P(1-K_P)$. Further, since X has only two values, the regression of Y on X must be linear, and the regression coefficient is $\mathrm{cov}_R/K_P(1-K_P)$, where cov_R is the covariance between relatives for the all-or-none trait, i.e. the covariance between X and Y. If K_R denotes the frequency of the attribute in relatives of index cases, then K_R is the mean value of Y at the point $X = 1$. Thus it can be calculated from the regression equation as

$$K_R = K_P + \frac{\mathrm{cov}_R}{K_P(1-K_P)}\,1-K_P \tag{1}$$

$$= K_P + \frac{\mathrm{cov}_R}{K_P}.$$

This general result involves no genetic theory. However, once a genetic model has been specified, the value of cov_R may be deduced. For each genotype the proportion of individuals with the trait is taken as the genotypic value and the variance in these values is partitioned into its components, V_A, V_D, V_{AA}, ..., in the standard manner (Kempthorne, 1957). Then if C_R is the chance that the relatives have a gene identical by descent at a locus, and u_R is the chance that they have two genes at a locus identical by descent,

$$\mathrm{cov}_R = C_R V_A + u_R V_D + C_R^2 V_{AA} + \ldots. \tag{2}$$

The regression equation may then be written as

$$K_R = K_P + \frac{V_A}{K_P} C_R + \frac{V_D}{K_P} u_R + \frac{V_{AA}}{K_P} C_R^2 + \ldots. \tag{3}$$

* On leave from the School of Wool and Pastoral Sciences, the University of New South Wales.

Under random mating the coefficients C_R and u_R have the following values for the most useful classes of relatives:

Class of relatives	Monozygotic twins	Parent–child	Full sibs	Half sibs	First cousins	Grandparent–grandchild
C_R	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$
u_R	1	0	$\frac{1}{4}$	0	0	0

As an illustration, let us suppose a trait is determined by two alleles A_1 and A_2 at a single locus A_2 having gene frequency q and mating being at random. The proportions of individuals having the trait are f_0, f_1 and f_2 among the genotypes A_1A_1, A_1A_2 and A_2A_2 respectively. Then the genetic parameters are (Kempthorne, 1957)

$$\left.\begin{aligned} K_P &= (1-q)^2 f_0 + 2q(1-q) f_1 + q^2 f_2, \\ V_A &= 2q(1-q)[q(f_2-f_1)+(1-q)(f_1-f_0)]^2, \\ V_D &= q^2(1-q)^2[f_2-2f_1+f_0]^2. \end{aligned}\right\} \tag{4}$$

Equations (1) to (4) allow the frequencies of a trait in relatives of index cases to be found directly for any two allele single locus model.

Two conclusions of interest are obvious from the present treatment. First, for a single locus model only three parameters, namely K_P, V_A and V_D, can be estimated from data on incidences in relatives, no matter how many classes of relatives are studied or how many parameters are included in the genetic model. This estimation problem may be regarded as one of fitting a multiple regression with K_R as dependent variate and C_R and u_R as independent variables. It can be seen from (4) that for any values of K_P, V_A and V_D one may take any value of q and solve the equations for f_0, f_1 and f_2. There are thus an infinite number of parameter sets $\{q, f_0, f_1, f_2\}$ which lead to the same frequencies in relatives. For example, both

$$\{q = 0{\cdot}2, f_0 = 0{\cdot}2, f_1 = 0{\cdot}3, f_2 = 0{\cdot}8\} \quad \text{and} \quad \{q = 0{\cdot}5, f_0 = 0{\cdot}176, f_1 = 0{\cdot}192, f_2 = 0{\cdot}464\}$$

give

$$K_P = 0{\cdot}256, V_A = 0{\cdot}010368, V_D = 0{\cdot}004096.$$

In our context only parameter sets for which f_0, f_1 and f_2 all lie between 0 and 1 are acceptable, so that no acceptable results may be obtainable for some gene frequencies. However, it is important that, if fitting is done in terms of q, f_0, f_1 and f_2, as by Elston & Campbell (1970), the possibility of fitting the data with a different set of parameters should be borne in mind.

The second conclusion concerns the problem of distinguishing between a single locus model and a multiple factor model such as that discussed by Falconer (1965) for disease liability. The essential difference between the models in our treatment is that single locus models cannot give rise to epistatic variance components. If the multiple factor hypothesis is true in any case it will give a better fit to frequencies in relatives only if it produces appreciable epistatic variance in the all-or-none trait. It was shown by Dempster & Lerner (1950) that epistatic variance is not large unless heritability of liability is high and the proportion of the population having the trait is near 0 or 1. But even if these necessary conditions for discrimination are satisfied, two further points must be borne in mind. The epistatic variance components have small coefficients in covariances between relatives, and these coefficients are correlated (among classes of relatives) with the coefficients C_R and u_R. Thus, although the models should be distinguishable in favourable material, it will often in practice be very difficult to make the distinction solely on the basis of frequencies in relatives.

REFERENCES

DEMPSTER, E. R. & LERNER, I. M. (1950). Heritability of threshold characters. *Genetics* **35**, 212–36.

ELSTON, R. C. & CAMPBELL, M. A. (1970). Schizophrenia: evidence for the major gene hypothesis. *Behavior Genetics* **1**, 3–10.

FALCONER, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics* **29**, 51–76.

KEMPTHORNE, O. (1957). *An Introduction to Genetic Statistics*. New York: John Wiley and Sons.

LI, C. C. & SACKS, L. (1954). The derivation of joint distribution and correlation between relatives by use of stochastic matrices. *Biometrics* **10**, 347–60.