Multifactorial Models for Familial Diseases in Man

By R. N. Curnow   and   Charles Smith†

Department of Applied Statistics,   *Department of Human Genetics,*
University of Reading   *Western General Hospital, Edinburgh*

[Read before the Royal Statistical Society on Wednesday, December 18th, 1974,
the President Professor H. E. Daniels in the Chair]

Summary

Some familial diseases may be caused by many factors, genetic and environmental, acting jointly. The value and limitations of multifactorial models that have been proposed for the inheritance of these diseases are discussed. Topics considered include the complicating effects of common familial environment; the calculation of recurrence risks; discrimination between different models of inheritance; the resolution of disease heterogeneity; the use of associated continuous measurements; and the effects of selection against genes increasing liability to disease.

Keywords: FAMILIAL; HERITABILITY; MULTIFACTORIAL; LIABILITY; GENETIC DISEASE; GENETIC HETEROGENEITY; THRESHOLD; RECURRENCE RISK

1. Introduction

Problems in assessing the importance of heredity in human disease and in the expression of normal traits in man have long concerned geneticists and biometricians alike. Indeed much of the early work in biometrical statistics stemmed from problems in human genetics. So it is perhaps fitting that we should discuss some recent applications of biometrics in the study of human disease to a meeting of statisticians.

Diseases with an appreciable genetic component in their causation become proportionally more important with the decline in the frequency of diseases caused mainly by infection or by poor environment and nutrition. The number of live-born children dying in the first year of life in England and Wales has fallen from 133 per 1,000 in 1902 to 22 per 1,000 in 1960 (see Carter, 1969). However, the number certified as dying from congenital malformations has remained over this period at about $4\frac{1}{2}$ per 1,000. About 20 children per 1,000 are born with a severe or moderately severe physical malformation. Many of these conditions show familial aggregation, that is there is an increased frequency in the relatives of affected individuals, and are probably partly genetic in origin. A World Health Organisation report (1972) found that 30 per cent of admissions to one North American paediatric hospital and 40 per cent of paediatric deaths in the United Kingdom were more or less directly related to "genetic" disease.

Diseases with an appreciable genetic component in their causation can often be prevented or treated as can most other forms of disease in the population. In families with a family history of a disease, further cases may be prevented by genetic counselling, or by antenatal diagnosis and selective abortion, or by special care of individuals born at risk. Population screening programmes can be applied to couples, pregnancies or to the newborn, to allow prevention or early detection. Understanding the

aetiology, including the form of inheritance, and the risks of genetic disease are important in work on prevention, treatment and care.

The object of this paper is to review multifactorial models of disease inheritance in man and to discuss their value, and their limitations, in theory and in practice. Section 2 gives a brief background review of diseases inherited in a simple Mendelian manner. Section 3 attempts to deal with the problem in familial diseases of distinguishing between the effects of genetic inheritance and the effects of common familial environment. Section 4 gives a formal statistical account of multifactorial models. These models are then developed and applied to various situations in the later sections of the paper.

With many authors contributing, the terminology used has become varied and often confusing, so some standardization is needed. A disease or disorder or condition, is said to show familial aggregation if the proportion affected is raised in relatives of affected individuals and genetic only if genetic effects are established. Prevalence refers to the proportion of cases existing in a population at a given time. Incidence deals with the proportion of new cases occurring in a given population over a given period, e.g. 1 year or a life time. A familial disease may be termed polygenic (if many genetic factors are proposed) or multifactorial (if many factors of unspecified type are proposed). A disease may be termed semi-continuous or quasi-continuous if it corresponds to the division of some underlying continuous scale into two all-or-none $(0, 1)$ classes corresponding to diseased and not diseased.

2. SIMPLY INHERITED DISORDERS

Familial diseases in man are usually divided into three main groups: (1) those due to a single genetic locus and usually inherited in a simple Mendelian manner, (2) those due to a known chromosomal abnormality and (3) other familial diseases where groups (1) and (2) have not so far been demonstrated. Some examples of Mendelian and chromosomal disorders are given in Table 1, showing the different common forms, with the genotype of affected individuals, and the most common mating type involved. The proportion of affected offspring produced, and the risk to each subsequent child, is $\frac{1}{2}$ or $\frac{1}{4}$ for simple Mendelian disorders but very much less for incompatibilities and for chromosomal disorders. There is a large number of simple Mendelian disorders and traits known in man (McKusick, 1971) and though individually rare, these disorders are cumulatively important. Dominants outnumber recessives, probably because it is easier to establish their mode of inheritance from family material. In other species where breeding experiments can be performed, such as the mouse, autosomal recessive forms of abnormality are the most frequent.

The basis for classifying a disease as Mendelian depends mainly on showing strict adherence to the simple $1:1$ or $1:3$ ratios expected of genes, x-linked or autosomal. However, it may be difficult, even for simply inherited disorders, to establish the true mode of inheritance because many factors may distort the simple Mendelian ratios. An extensive methodology (Morton, 1969; Morton *et al.*, 1971) has been developed to estimate segregation ratios and other parameters. Many of the conditions are rare and data are hard to collect. There are problems of ascertainment of probands (index cases) and their families; of new mutations occurring; of errors in clinical diagnosis; of phenocopies (a similar clinical form of non-genetic origin); of illegitimacy; of variable family size; and of possible genetic heterogeneity (one clinical condition arising from different genetic loci).

TABLE 1

Mendelian and chromosomal forms of genetic disease

Mode of inheritance	Abnormal genotype	Common parental mating type	Risk to the next child (segregation ratio)	Typical disorders	Number of disorders (traits) identified (McKusick, 1971)
Mendelian disorder					
Autosomal dominant ("single dose")	a*a	a*a × aa	$\frac{1}{2}$	Achondroplasia	415[2] (528)[3]
Autosomal recessive ("double dose")	a*a*	a*a × a*a	$\frac{1}{4}$	Phenylketonuria	365 (418)
X-linked recessive	x* (♂)	x × x*x (♀)	$\frac{1}{4}$	Haemophilia	86 (64)
Maternal–fetal incompatibility	Dd	D–Xdd (♀)	< 5%	Rhesus haemolytic disease	—
Chromosomal abnormality					
Sex chromosome	XO	Normal	< 1%	Turner's syndrome	—
	XXY	Normal	< 1%	Klinefelter's syndrome	—
Autosome	Trisomy	Normal	< 3%	Down's syndrome	—

* Indicates the abnormal allele; 2—probable; 3—additional possible.

The spectrum of familial disease from simple Mendelian forms to those with complex familial patterns is well illustrated in Fig. 1 (Newcombe, 1964), which plots the frequency in sibs of affected individuals against the frequency in the general population. The dominant and recessive autosomal disorders stand out from other
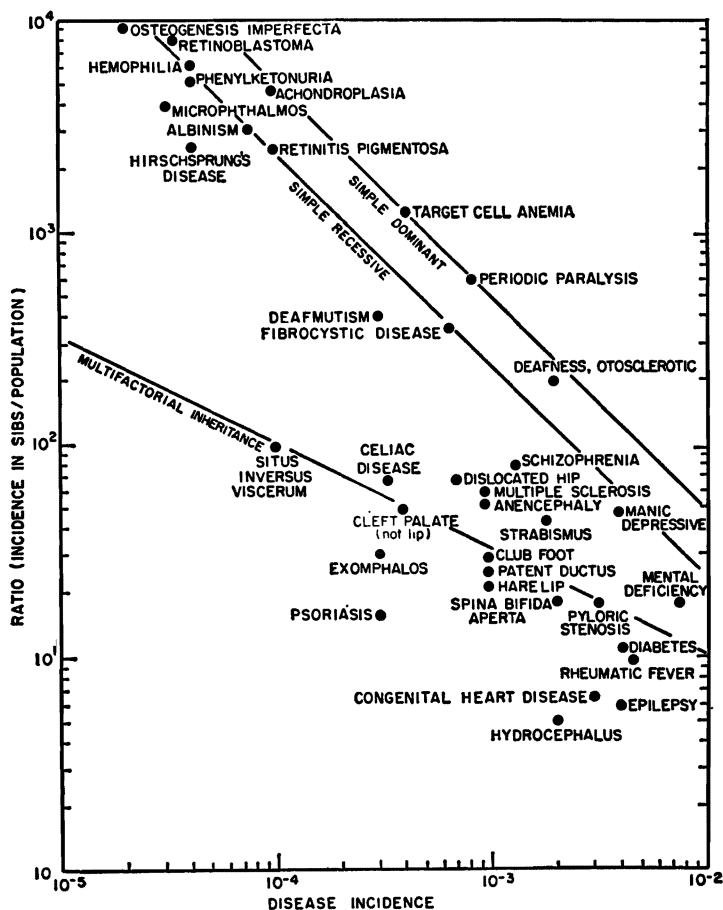

FIG. 1. Relation between disease incidence and relative incidence in sibs of affected individuals for a number of diseases. The lines indicate the expected relationships for simple dominant, simple recessive and Edwards' (1963) approximation to multifactorial inheritance (from Newcombe, 1964).

familial disorders because of the high risk in sibs. Other diseases show features that cannot be readily explained by simple strict Mendelian inheritance at a single di-allelic locus. These features are:

(i) the high frequency and severity of the disease which together would imply either an unusually high mutation rate or some selective advantage of the heterozygote form (see Section 6);

(ii) the frequency of the disease in sibs (and other relatives) of affected individuals is lower than would be expected from single locus di-allelic inheritance; and

(iii) the frequency in relatives increases as the severity and as the number of family members affected increases.

Can the effects of environmental factors added to the single locus model explain the observed familial aggregation patterns of these diseases, or do we need a multifactorial model that involves the effects of genes at more than one locus?

The single locus model can be extended to include situations where the genes at the locus do not completely determine the presence or the absence of the disease but do influence the probability that the individual will have the disease (see Section 7). These generalized single locus models, which include simple dominance and recessive inheritance as special cases, can allow for various "noise" factors in the expression of the disease, such as variable penetrance, errors in clinical diagnosis, phenocopies, genetic heterogeneity, the variable manifestation of heterozygotes and the sporadic non-heritable occurrence of the disease. An extension of the model with correlations between relatives in this variable translation of genotype into disease would allow the incomplete penetrance to be due to genes at other loci or to environmental factors common to relatives. This would be a multifactorial model but with one locus having a major effect on liability. We shall see later that there are difficulties in discriminating between single locus models and multifactorial models representing the small independent effects of many loci and many environmental factors. For this reason, and because few examples have been established, intermediate models involving two or three genetic loci will not be considered.

3. COMMON FAMILIAL ENVIRONMENT

Once we allow environment to be a causative factor in a disease, we have to recognize the fact that members of the same family tend to share the same environment as well as the same genes. Some of the correlations in disease incidence among family members may be due to common familial environments. This problem has long plagued studies of heredity in man since the environment cannot be randomized among individuals and so the effects of common family environment and common genes are confounded. Attempts to resolve this problem should be the first step in any study of familial disease in man, and yet it is very frequently ignored. Unless common familial environmental effects can be discounted or adjusted for, then estimates of any genetic parameters may be seriously in error and misleading.

The main method used to measure the importance of common familial environments is to include in the study unrelated individuals living together. This could include adopted children and their adoptive parents and adoptive sibs; spouses; relatives' spouses; and individuals in institutions. The assumption is made that choice in adoption or marriage is not directly or indirectly associated with the disease or trait being studied.

Another way to avoid the complications of common familial environments is to estimate genetic parameters from related individuals living apart. These could be adopted children and their natural parents. Alternatively, differences in relationship of related individuals living together, such as monozygotic and dizygotic twin pairs, can be used. Usually, however, the data available on such groups are limited and special searches need to be made.

A good example of the use of such materials in assessing the importance of common familial environmental effects is given for schizophrenia by Rosenthal (1970) and summarized in Table 2. Psychiatric diseases generally, and schizophrenia in particular

may be influenced by complex social relations within families. For this reason, both biological and adoptive relatives of adopted schizophrenics and of adopted controls were studied. The frequency of schizophrenia in the biological relatives of the schizophrenia patients was very much higher than in their adoptive relatives when compared with the control groups. This and other similar studies (Heston, 1966;

TABLE 2

The frequency of schizophrenia in biological (first degree) and adoptive relatives of adopted schizophrenics and adopted controls (from Rosenthal, 1970)

Adopted index cases	Biological relatives			Adopted relatives		
No.	No.	Affected	Affected (%)	No.	Affected	Affected (%)
33 Schizophrenia	150	13	8·7	74	2	2·7
33 Control	156	3	1·9	83	3	3·6

Schulsinger, 1972) suggest that genetics is an important factor in causing the familial aggregation commonly found in schizophrenia and other psychiatric disorders. However, it must be remembered that the closeness of a biological relationship as opposed to an adoptive one is not purely a matter of genes shared.

4. MULTIFACTORIAL MODELS

4.1. *The Liability Model*

In one standard multifactorial model some single-dimensional quantity x, called liability, is assumed to determine the probability of an individual succumbing to the disease. The x-values may be determined both by genes and by environment. The x-value may, for example, be concerned with a development rate in congenital malformations such as spina bifida, the concentration of some biochemical product in a metabolic disease or blood pressure in hypertensive disease. All correlations in the occurrence of the disease in relatives are induced by correlations between the x-values of the relatives. If liability is determined by many genetic or environmental factors the values of x in the population may be assumed to have a continuous distribution. We can then, in theory, and without loss of generality, transform x so that it has a Normal distribution over the population with mean zero and unit variance. There is a real assumption when we assume that, in addition to the Normality of the marginal distributions, the joint distribution of the x-values for k-relatives has a k-variate standard Normal distribution. This will be assumed in all that follows.

Let $S(x)$ denote the probability that an individual with value x succumbs to the disease and $f_k(x_1, x_2, ..., x_k; \rho)$ denote the joint density function of the standardized k-variate Normal distribution with correlation matrix ρ. In the univariate case, we shall write

$$f_1(x_1) = \phi(x_1)$$

and

$$\int_{-\infty}^{X_1} f_1(u)\, du = \int_{-\infty}^{X_1} \phi(u)\, du = \Phi(X_1).$$

The frequency of the disease in the population will be

$$P_1 = \int_{-\infty}^{+\infty} \phi(u)\, S(u)\, du$$

and the probability that all k relatives in a family have the disease is

$$P_k = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_k(u_1, u_2, \ldots, u_k;\, \rho)\, S(u_1)\, S(u_2) \ldots S(u_k)\, du_1\, du_2 \ldots du_k.$$

If we can calculate the probability of each of all possible sub-sets of relatives all having the disease, then we can calculate, by simple probability arguments, the probability of any sub-set having the disease and the rest not having the disease.

One form suggested for $S(x)$ is (Edwards, 1969)

$$S(x) = a\, e^{bx} (a, b > 0,\ -\infty < x + \infty).$$

This form simplifies the analysis since the density function of x among the affected members of the population is

$$\phi(x)\, S(x) \Big/ \int_{-\infty}^{+\infty} \phi(x)\, S(x)\, dx = \phi(x - b).$$

Thus the distribution of x among the affected individuals is the same as among the whole population but with the mean increased by an amount b. Unfortunately, the risk $S(x) = a\, e^{bx}$ does exceed one for sufficiently large x. Edwards' argument that $S(x)$ will only be greater than one for rare values of x is insufficient to justify results that may well be heavily influenced by rare "probabilities" appreciably exceeding one.

Great simplifications in the necessary computations result if we can assume instead that $S(x)$ is a sigmoid function. We shall assume this and write

$$S(x) = \Phi\!\left(\frac{x - \mu}{\sigma}\right).$$

The sigmoid function is an appropriate risk function since it increases monotonically from 0 to 1 as x increases from $-\infty$ to $+\infty$. We could allow for a base-line incidence of the disease, α, and an uncertainty of disease, $\beta < 1$, when x is $+\infty$ by defining

$$S(x) = \alpha + (\beta - \alpha)\, \Phi\!\left(\frac{x - \mu}{\sigma}\right).$$

In practical situations, this would require the estimation of the parameters α and β and this may well prove difficult.

We shall assume that $S(x) = \Phi\{(x - \mu)/\sigma\}$. The incidence of the disease in the population will be

$$P = \int_{-\infty}^{+\infty} \phi(x)\, \Phi\!\left(\frac{x - \mu}{\sigma}\right) dx.$$

We may interpret this probability as

$$P = \mathrm{Prob}\!\left(z < \frac{x - \mu}{\sigma}\right),$$

where z and x are independent standardized Normal variables. Therefore,

$$P = \text{Prob}\left(\frac{\sigma z - x}{\sqrt{(\sigma^2+1)}} < -\frac{\mu}{\sqrt{(\sigma^2+1)}}\right)$$

$$= \Phi\left(-\frac{\mu}{\sqrt{(\sigma^2+1)}}\right).$$

Clearly, $\theta = -\mu/\sqrt{(\sigma^2+1)}$ determines the population incidence of the disease and σ the sensitivity of the probability of disease to changes in the value of x.

The probability that all of k relatives succumb to the disease is

$$P_k = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_k(x_1, \ldots, x_k; \rho) \Phi\left(\frac{x_1-\mu}{\sigma}\right) \cdots \Phi\left(\frac{x_k-\mu}{\sigma}\right) dx_1 \ldots dx_k$$

$$= \text{Prob}\left(z_1 < \frac{x_1-\mu}{\sigma}, \ldots, z_k < \frac{x_k-\mu}{\sigma}\right),$$

where $z_1, z_2, \ldots, z_k$ are independent standard Normal variables distributed independently of the jointly Normally distributed variables $x_1, x_2, \ldots, x_k$. Therefore,

$$P_k = \text{Prob}\left(\frac{\sigma z_1 - x_1}{\sqrt{(\sigma^2+1)}} < -\theta, \ldots, \frac{\sigma z_k - x_k}{\sqrt{(\sigma^2+1)}} < -\theta\right)$$

$$= \int_{-\infty}^{-\theta} \cdots \int_{-\infty}^{-\theta} f_k(t_1, t_2, \ldots, t_k; \rho^*) dt_1 \ldots dt_k,$$

when

$$\theta = \frac{\mu}{\sqrt{(\sigma^2+1)}} \quad \text{and} \quad \rho^* = \frac{1}{\sigma^2+1}\rho.$$

The multiple integrals in P_k can be reduced to single integrals involving univariate Normal density and cumulative distribution functions if the matrix ρ takes particular simple forms (see Curnow and Dunnett, 1962, for the general method and references).

4.2. *Equivalent and Alternative Models*

The liability and risk function approach outlined in the previous section is mathematically equivalent to the abrupt threshold model (Falconer, 1965, 1967) which has been most commonly used in practice. Earlier workers with threshold models include Pearson (1900, 1904, 1914), Wright (1934), Robertson and Lerner (1949), Dempster and Lerner (1950), Gruneberg (1952) and Crittenden (1961). In Falconer's model, a Normally distributed quantity z with variance σ^2 is added to the underlying x-value for an individual. Then all individuals with values of $y = (x+z)$ greater than a certain threshold, μ (Falconer's T), manifest the disease while those with y less than μ do not. $S(x)$ is then the probability that $z > (\mu - x)$ and is therefore $\Phi\{(x-\mu)/\sigma\}$. The larger the value of the threshold, μ, the lower the frequency of the disease in the population. There is some arbitrariness in the division between x and z when, as is assumed here, x is Normally distributed and the $S(x)$ function sigmoid. All that matters is that the z-values for different individuals must be independent. Although the mathematics is the same, the idea of an abrupt threshold is less acceptable biologically than the idea of a risk function (Edwards, 1969; Smith, 1970, 1971a).

In some of the earlier work on threshold models, approximations were involved concerning the distribution of y, $= x + z$, among relatives of individuals known to be affected. This distribution was assumed to be Normal with a different mean but the same variance as the distribution for the whole population. Edwards (1969) showed how these approximations could be avoided when disease information was available on a single relative by using singly truncated forms of the bivariate Normal distribution. Aitken (1934) derived expressions for the means, variances and covariances of jointly Normally distributed variables following truncation based on a sub-set of these variables. Mendell and Elston (1974) and Reich *et al.* (1972) used these expressions to derive frequencies in relatives and hence to obtain approximate risk estimates when disease information was available on more than one relative. These latter estimates were still only approximate because, although the means, variances and covariances were now correct, the Normality of the y distribution for relatives was assumed still to hold despite the truncation exercised on the correlated y-values of affected individuals. The method due to Smith (1970) to be described in Section 5.1 and that due to Curnow discussed earlier in this section have removed this final approximation from the calculation of risks.

To interpret the above results about the threshold model into the form commonly used in quantitative genetics, we need to assume that all of the causes of familial correlation are genetic, or that any non-genetic familial effects have been removed. Then assuming that the genetic variance is entirely additive the heritability (Falconer, 1965) of liability ($y = x + z$) can be defined, in the present notation, as

$$h^2 = \frac{\mathrm{var}\,(x)}{\mathrm{var}\,(x+z)} = \frac{1}{\sigma^2 + 1}.$$

The correlation in liability between any pair of relatives is then

$$\rho^* = \rho/(\sigma^2 + 1),$$

where ρ is the genetic relationship between the relatives concerned (see Section 5.1). The threshold model can be used when the genetic variance includes dominance and epistatic components providing only that the correlations of liabilities are estimated from the appropriate types of relatives or from a knowledge of the values of the various components of the genetic variance in the population studied. Care will have to be taken if there are correlations between the liabilities of parents.

In much of the work now to be described the following assumptions are made:

   (i) that there are no birth-order effects,

   (ii) that the x-values of parents are uncorrelated, and

   (iii) that the correlations between a child and a parent and between two sibs are equal.

The last assumption would be true if the variance in x was entirely additively genetic and the environmental and infective correlations between sibs and between parents and offspring were all equal. The predictions made for other relatives require stricter assumptions such as that the correlations are entirely due to genetic effects and to the additively genetic variance.

Morton *et al.* (1970) have suggested an alternative multifactorial model for disease inheritance based on the concept of genetic load. If the risks $p_1, p_2, ..., p_n$ arising from each of a large number, n, of factors are small and independent, then the total risk to

an individual can be written

$$P = 1 - \prod_{i=i}^{n} (1-p_i),$$

and approximated by the formula

$$P = 1 - \exp\left(-\sum_{i=i}^{n} p_i\right).$$

$A = \sum_{i=i}^{n} p_i$ is called the load. Sex differences and the effects of inbreeding can be allowed for but, without these complexities, the load for a relative with degree of relationship R is $A + CR$ where A is the population frequency and C is a measure of the additive effects summed over loci. Given the population frequency and the frequencies in relatives of affected individuals, values of A and C can be estimated. The model does allow dominance and, providing the exponential approximation is still sufficiently accurate, large effects on the load scale. This model will not be discussed further. In general it provides results very similar to those derived from the multifactorial model presented in this section and the single locus models with incomplete penetrance to be developed in Section 7. The only exceptions to this similarity occur when the penetrance is virtually complete or where the frequency of the disease is low and the heritability of the multifactorial model is high.

5. APPLICATIONS OF THE MODEL

5.1. *Assumptions and Recurrence Risk Calculations*

Much of the application of the multifactorial model so far has been concerned with the estimation of $\rho^* = \rho/(1+\sigma^2)$, from information about the incidence (P_1) of the disease in the population and the incidence P_2/P_1, of the disease in relatives of affected individuals. We shall not discuss the detailed problems of estimation (see Reich *et al.*, 1972; Draper 1974; Mendell and Elston 1974). The relatives most usually considered have been first-degree relatives (parents, children, brothers or sisters) but monozygotic twins, second-degree relatives (uncles, aunts, nephews, nieces and grandparents) and third-degree relatives (cousins) have also been studied. If the correlations between the x-values of relatives are due to additive genetic effects at many loci, then $\rho = 1$ for monozygotic twins, $\rho = \frac{1}{2}$ for first-degree, $\rho = \frac{1}{4}$ for second-degree and $\rho = \frac{1}{8}$ for third-degree relatives. μ and σ can be estimated given estimates of P_1 and ρ^*. The single integral forms for P_k mentioned above can then be used to derive probabilities of disease patterns for some simple groups of relatives, and hence recurrence risks for individuals with particular family histories of disease. The probabilities can be used to check the adequacy of the model using information available on the frequency of familial patterns of disease and the recurrence risks can be used in genetic counselling. Curnow (1972) used the integral reductions to tabulate the risks for individuals given disease information on their parents, on one parent and one sib; or on one, two or three sibs. He also tabulated risks given information on a monozygotic twin; two such twins; or on a twin and a sib or a parent.

Smith (1971a) used numerical integration to study a wide range of family situations. Dividing the range of values of the underlying quantity into a large number of small non-overlapping intervals (i), the frequency (f_i), the probability (P_i) that an individual with underlying value at the midpoint of the interval succumbs to the disease, and the probability (P_i') that a particular relative (e.g. a sib) succumbs were

derived. The frequency of the disease among these relatives, that is the recurrence risk, is then

$$\sum_i f_i P_i P_i' \Big/ \sum_i f_i P_i,$$

where the summation is over the intervals. By increasing the number of intervals any desired accuracy can be achieved. Graphs were then drawn, for various values of $\rho*$, of the frequency of the disease in relatives against the frequency (P_1) in the population. Recurrence risks in individual families given varying amounts of information, both positive and negative, about members of the family were also studied. The distributions of x for all original and all intermediate members of independent branches of the family were split into classes. The probabilities of patterns of occurrence of the disease for given sets of classes can then be added over all combinations of classes, weighted by their frequencies to obtain the recurrence risks. The method can also take account of sex and severity differences, and differences in heritability with age. For the more complex family histories approximate methods must be used (Smith 1971a). Smith also derived confidence limits for the recurrence risks. Fig. 2 shows some of the results obtained by Curnow (1972) and Smith (1970, 1971a). The recurrence risks and population frequencies are on logarithmic scales. With these scales the recurrence
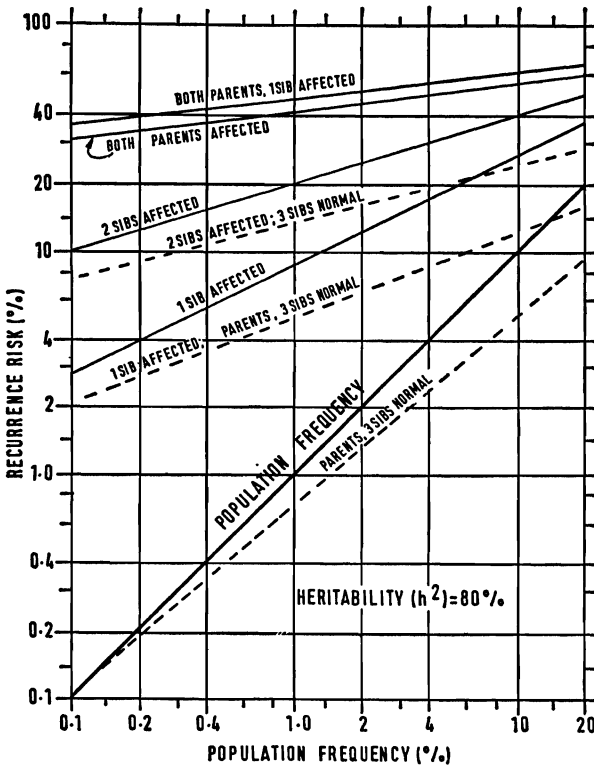

FIG. 2. Recurrence risks for multifactorial inheritance in a variety of sibships, with heritability (h^2) of 80 per cent. Broken lines include normal relatives.

risks are approximately linearly related to population frequency allowing easy inter-polation. With two or more first-degree relatives affected the risks increase sub-stantially. The inclusion of unaffected relatives decreases the risk only slightly and they may be ignored, unless the disease incidence is high.

For simple cases, for example with one relative affected, observed or "empiric" risks are available and the risks are usually low—less than 5 per cent. These risk estimates are termed empiric, because they do not depend on any genetic model. However, they are average values and do not allow for differences in risk between families dependent on their detailed family history, in terms of severity, age of onset and sex. A range of risk estimates may be needed depending on the age and sex of the individual to whom the risk estimate is to be applied. When such factors have to be considered or when there are two or more affected individuals in the family there are usually no empiric risk estimates available for use in counselling.

Further information on the family history, including information about second- and third-degree relatives, is usually available. Estimation of risks can thus become very complex and the risk may need to be evaluated uniquely for each particular family. A computer program (RISKMF) is available to do this (Smith, 1972). It takes all the above factors into account, but only approximately for second- and third-degree relatives. A series of risk tables has also been prepared for some 180 possible family histories for the major congenital abnormalities (Bonaiti-Pellié and Smith, 1974) and these may be useful in genetic counselling.

TABLE 3

Estimates of correlation in liability for some congenital abnormalities and for schizophrenia

Some congenital abnormalities	Population frequency (%)	Frequency in relatives (%)	Correlation in liability	Author
Cleft lip ± cleft palate (first-degree relatives)	0·1	3·1	0·41 ± 0·02	Carter (1969)
Spina bifida and anencephaly (sibs)	0·29	4·4	0·38 ± 0·02	Carter and Evans (1973)
Congenital pyloric stenosis (first-degree relatives)	0·30	4·0	0·37 ± 0·02	Carter and Evans (1969)
Schizophrenia	1·00			Gottesman and Shields (1967)
MZ twins		53	0·89 ± 0·06	
DZ twins		14	0·46 ± 0·05	
Sibs		10	0·39 ± 0·02	
All first-degree relatives		10	0·39 ± 0·02	
All second-degree relatives		4·6	0·28 ± 0·02	

As a simple example of the calculation of liability correlations from which com-plex risks could be derived, consider the condition cleft lip with or without cleft palate (Carter, 1969). The incidence of this congenital abnormality is about 1 per 1,000 and about 31 per 1,000 in sibs of affected individuals. The estimated correlation in liability among sibs from this information is $\rho^* = 0.41 \pm 0.02$. Correlation coefficients for several other congenital abnormalities are given in Table 3 (Bonaiti-Pellié and Smith,

1974). Correlation estimates can be derived from different groups of relatives as is shown for schizophrenia in Table 3. To compare the different estimates they have to be converted into estimates of heritability of liability, $1/(1+\sigma^2)$, by taking account of the degree of relationship, ρ, between the relatives and by removing or discounting any common familial environmental effects contributing to the correlation. Significant differences among heritability estimates from different relatives, or heritabilities exceeding unity may suggest that dominance or epistasis may be important, that the multifactorial model may not apply or that non-genetic familial effects have not been adequately discounted.

5.2. *Sex and Age Effects and Differences in Severity*

In estimating the correlation between the liability, x, of relatives, there are often complications because the disease may occur with varying severity and the disease frequency may depend on age and sex. In some disorders the sex with the lower incidence often has the higher frequency of affected relatives. This apparent reversal of frequencies is quite consistent with the multifactorial model. This is because the sex with the lower frequency will have its risk function $S(x)$, $= \Phi\{(x-\mu)/\sigma\}$, displaced to the right of the risk function for the other sex. Hence, an affected individual of the sex with the lower frequency will probably have a higher x, or liability value, than affected individuals of the other sex. Their relatives will also tend to have higher liabilities and so a higher proportion of them will be affected. A good example of this is given by the congenital abnormality pyloric stenosis, in Table 4. Females are less frequently affected but their relatives are at higher risk and the pattern appears confusing. However, when the correlations are estimated from the different sets of relatives (comparing the frequency in relatives with the population incidence in the same sex) the anomaly in the frequencies is largely resolved and the estimates can be pooled to give a single estimate of liability correlation for the disorder.

TABLE 4

Analysis by sex of familial frequencies in pyloric stenosis (after Falconer, 1965)

		First-degree relatives				
		Male			Female	
Proband	Population frequency (%)	Frequency (%)	Correlation in liability		Frequency (%)	Correlation in liability
Male	0·5	5·0	0·32±0·04		2·2	0·37±0·06
Female	0·1	17·1	0·48±0·05		6·6	0·47±0·07

Pooled estimate 0·40±0·025.

To take account of different levels of severity of a disorder two, or more, risk functions differing in location, i.e. in μ, can be used corresponding to the different severity classes. Similarly, the onset age of a disorder, such as diabetes, may be associated with liability and a range of risk functions may be constructed for the different age groups.

Moving the risk function, $\Phi\{(x-\mu)/\sigma\}$, relative to the mean of the liability distribution changes only μ, not σ. A further possibility would be to allow differences in σ, and hence in heritability for the two sexes, for different age groups and for different severity classifications.

Several other factors may be more difficult to take into account in analysis, such as (1) differential mortality of affected individuals (Smith *et al.*, 1972; Draper, 1974), (2) inappropriate or unreliable estimates of incidence (Smith, 1974) and (3) different rates of detection and diagnosis for patients and for relatives (Smith, 1974).

5.3. *Concordance in Twins*

Monozygous (MZ) twins have the same genotype, so if a condition is entirely genetic in origin, concordance in MZ twin pairs should be complete, as it is for the strict Mendelian disorders. This is rarely the case in the common familial diseases. Often the MZ concordance rates are quite low, even for conditions which would otherwise be thought to have an important genetic component. Here the concordance rate is the proband concordance rate, i.e. the proportion of affected co-twins of independently ascertained affected individuals. In fact low concordance rates are expected with the multifactorial (MF) model, especially in conditions with low population prevalence, as shown in Fig. 3 (Smith, 1970). This is because the prior risk to any individual with a given "genotypic" liability, is low and so, assuming no environmental correlation, the risk to an MZ co-twin with the same genotype as his affected twin is also low.
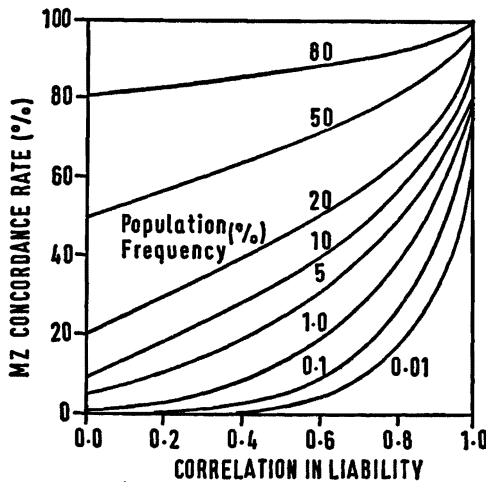

FIG. 3. Expected "proband" concordance rate in monozygotic (MZ) twins given the population frequency and the correlation in liability (from Smith, 1970).

This result may remove some of the confusion in comparing results in twins and other relatives. For example, estimates, from relatives other than monozygotic twins, of the heritability of clubfoot (talipes equinovarus) range from about 0·60 (Wynne-Davies, 1970) to 0·68 (Ching *et al.*, 1969). However, these estimates seemed to conflict with Idelberger's (1939) result of a 33 per cent concordance rate for monozygotic twins. Yet, with a population incidence in Caucasians of 0·12 per cent,

Fig. 3 shows that a concordance rate of 33 per cent is in fact not too low but instead is rather high, for heritabilities in the range 0·60–0·68.

5.4. *Resolution of Genetic Heterogeneity*

The frequency of one disease in relatives of patients with another disease can be used to measure the degree of genetic association between two diseases or to resolve genetic heterogeneity. This procedure is, of course, used intuitively by physicians in grouping or resolving various clinical forms or groups of disease. A simple 2×2 χ^2 test of the numbers with form 1 and with form 2 among relatives of patients with form 1 and of patients with form 2 can be used to test for complete association. For example, spina bifida and anencephaly are usually classified as different abnormalities but they run together in families as shown from data of Carter *et al.* (1969) in Table 5. The 2×2 χ^2 of the numbers in Table 5 was not significant and, with similar results from other studies, the two forms are usually considered as different manifestations of the same genetic disorder.

The same procedure can be elaborated and quantified using the multifactorial model of disease liability. Given two groups—separated on any criterion, clinical, biochemical or statistical—the genetic correlation in liability (Falconer, 1967) can be estimated as

$$h^2_{12}/\sqrt{(h^2_{11} h^2_{22})} \quad \text{or} \quad h^2_{21}/\sqrt{(h^2_{11} h^2_{22})},$$

where h^2 is heritability, the first subscript refers to patients and the second to relatives, and 1 and 2 refer to the two disease groups separated. For example, in Table 5 the estimates of genetic correlation between spina bifida and anencephaly are very high, showing that the two conditions are closely associated and can be treated in risk estimation and in genetic analysis as one condition. Similarly to test for genetic independence, a null hypothesis of a genetic correlation of zero can be tested. Note that if a disease is made up of two or more independent sub-groups, then the heritability of the combined condition will be decreased (not increased as suggested by Edwards, 1969).

Falconer's (1967) simple method for estimating the genetic correlation in liability between two diseases or groups depends on assuming that the genetic correlation is zero, or is unity (Smith *et al.*, 1972). If it is intermediate then the method is not strictly appropriate due to overlapping of the distributions of the two forms. However, it has been found (Smith, unpublished) that the simple estimates of the genetic correlation will give a very good indication of the true genetic association between the two groups and is unlikely to mislead the investigator in interpreting his data.

5.5. *Associated Continuous Traits*

So far, estimates of recurrence risks have used only the disease status, normal or affected, of relatives. Often there is additional information on graded or continuous traits associated with the condition that can be incorporated to improve the estimated risk in a particular family. For example, blood pressure in hypertension, intra-ocular pressure in glaucoma or blood glucose levels in diabetes could be measured in the person at risk and in relatives and these would be informative in estimating recurrence risks for these diseases. The trait may define the disease, be a factor in its causation or be a result of the disease. Alternatively, the trait might be a measure of environmental factors involved.

TABLE 5

Genetic association of anencephaly and spina bifida (data from Carter et al., 1969)

Patients	Population frequency (%)	Total No.	Sibs								
			Anencephaly			Spina bifida			Combined		
			No.	Frequency (%)	h^2_{ij}	No.	Frequency (%)	h^2_{ij}	No.	Frequency (%)	h^2
Anencephaly	0·36	707	16	2·26	0·46 ± 0·07	13	1·84	0·36 ± 0·06			
Spina bifida	0·42	854	20	2·34	0·48 ± 0·06	32	3·75	0·58 ± 0·04			
Combined	0·78	1561							81	5·19	0·58 ± 0·03

Genetic correlation estimates

Patient	Relative	Genetic correlation $\{h^2_{ij}/\sqrt{(h^2_{ii}.h^2_{jj})}\}$
Anencephaly	Spina bifida	0·71 ± 0·20
Spina bifida	Anencephaly	0·92 ± 0·19

The important assumption that will be made here is that the risk of the disease is related to the continuous trait only through the latter's correlation with the liability, x. The effect of knowing the value of the correlated trait in the individual or in some relatives is then simply to change the mean values of the liabilities and the variances and covariances of these liabilities. As before, Aitken's (1934) formulae for the adjustment of variances and covariances to allow for truncation on these variables can be used to obtain approximate risk values (Smith and Mendell, 1974). The approximation again being that truncation on other variables has affected means, variances and covariances but not the Normality of the distributions. Curnow's exact method based on reduction of risks to single integral forms can be used when information is available on the individual and one relative and to more extensive situations providing the pattern of the liability correlation matrix takes the required structural form (Curnow, 1974). The approximate results (Smith and Mendell, 1974) agree well with the exact results derived by Curnow (1974) when only one relative is involved but the accuracy when several relatives are involved is not yet known.

The mean, variance and heritability of the correlated trait can be estimated from data in the usual way. The correlation between the trait and liability is estimated from a comparison of the mean values of the trait for affected and for unaffected individuals.

The calculations made show that estimates of risk can be substantially changed by inclusion of data on the associated trait. The overall value of the associated trait in increasing the precision of the risk estimate will depend on its correlation with liability to the disease. A good measure of its value is the additional proportion of the variation in liability accounted for by the associated trait, above that explained by the information on disease status (Smith and Mendell, 1974). If the correlation with liability is low then the trait adds little information and if the correlation is high then the individual's own value for the trait (as one's own blood sugar level in diabetes) gives a better estimate of liability to the disease than all the family history. Thus it is largely when the correlation between the trait and liability is intermediate, or when the individual at risk cannot be measured, perhaps because he is too young or not yet born, that an associated trait will be used with family history in risk estimation. A computer program (RISKCT) is available to do the necessary calculations.

Hopefully, clinical research will lead us to a fuller understanding of the nature of the liability, x. The results obtained about the relevance of a correlated trait include as special cases the trait being x itself or being an estimate of x subject to error. We are therefore able to test hypotheses that particular measurable quantities are x or estimates of x. It must be remembered that we have not assumed that x determines the occurrence of the disease, but only that x is a partial determinant of the disease that includes all the factors leading to correlations between the incidence of the disease in relatives.

If the underlying variable is identified, then a series of thresholds can be selected and estimates of the correlation in liability between relatives can be derived for each threshold level. Reich *et al.* (1972) found good agreement in estimates of the correlation for data on the number of lung tumours in mice exposed to urethane. However, Trimble (1971) working with some half a millon records on birth weight in man found highly significant differences between estimates from thresholds at different parts of the distribution. He was unable to get a transformation to a Normal distribution and concluded that the methods may be very sensitive to departures from Normality and warned against uncritical application of the models.

6. DISEASE FREQUENCIES AND SELECTION

Natural selection is continually reducing the frequency with which individuals with certain diseases reproduce. This will lower the frequency of any genes that increase the liability to these diseases. It is therefore reasonable to ask why so many diseases with a genetic component in their causation have not been eliminated from the population.

Severe recessive diseases tend towards an equilibrium in which the incidence of the disease is equal to the rate of mutation from the normal to the harmful form of the gene at the locus concerned. Mutation rates are thought to be of the order of 10^{-5} or much less (Cavalli-Sforza and Bodmer, 1971) so the balance between natural selection and mutation may explain the frequency of many of the recessive inborn errors of metabolism such as phenylketonuria, with its frequency of about 7×10^{-5}. The selection pressure against one of many genes acting additively, or otherwise, on liability in a multifactorial model will be less than the pressure at a single locus fully determining the occurrence of the disease. This will result in higher equilibrium frequencies for the harmful genes.

Another possible mechanism for maintaining deleterious genes in populations is a selective advantage for heterozygotes over both homozygotes. The best known example of this is sickle cell anaemia which is caused by a recessive allele. Heterozygotes for the sickling allele have a fitness some 25 per cent superior to that of normal homozygotes in areas where malaria is endemic. The recessive homozygotes have very low fitness and the frequency of the diseases in malaria areas is about 10 per cent. Similar, but unknown, mechanisms have been proposed to account for the high frequency of some of the inborn errors of metabolism such as cystic fibrosis in Caucasians. A 2 per cent advantage in fitness for heterozygotes over the normal homozygotes is almost sufficient to account for the current frequency of the disorder (5×10^{-4}) but an advantage of only 2 per cent would be very difficult to detect in practice, particularly since heterozygotes cannot, as yet, be reliably distinguished from normal homozygotes. Individual loci in the multifactorial model may also have genes held in equilibrium by similar forces.

Genes could also be held in equilibrium if individuals with a high liability value but not suffering from the disease had a slight superiority in fitness compared with individuals with a low liability value. A reduced fitness for individuals with low liability values would result in the preferential selection of individuals with intermediate liability values. This could lead to equilibrium gene frequencies but doubts exist about the stability of such equilibria (Robertson, 1956; Curnow, 1964).

So far in discussing selection we have assumed that large populations are involved. An abnormally high frequency of some Mendelian and multifactorial disorders in small isolates and in some populations may be due to the founder effect or to the random drift of gene frequencies (Rao and Morton, 1973).

Even with large populations, the relevance of equilibrium results depends on a constant environmental and genetical background for selection and mutation over many generations. The changes in gene frequency each generation are often small, of the order of the mutation rate or the selective pressures at individual loci. We may therefore be observing the effects of genes that previously had selective advantages or disadvantages but are now moving slowly towards new equilibria or towards elimination. For example, consider a gene that was previously at equilibrium as a recessive for a lethal disease but the disease has been harmless for the last t generations.

The gene frequency will be

$$q_t = 1 - (1-u)^t(1-u^{\frac{1}{2}}).$$

With $u = 10^{-5}$, it takes 130 generations to multiply the gene frequency by $\sqrt{2}$ and hence double the frequency of the now harmless disease.

7. DISCRIMINATION BETWEEN DIFFERENT MODES OF INHERITANCE

Is it possible to discriminate between the different modes of inheritance proposed for familial diseases using data likely to be available? Edwards (1960) in a now classic paper showed that it would be very difficult to discriminate between different modes of inheritance. All models tended to give similar familial patterns of frequencies so that differences in, for example, the predicted fall-off of incidence with decreasing genetic relationships, would be slight.

Moreover, a great variety of genetic models could be proposed for testing. Work in this area has tended to try to discriminate between a single-locus, two-allele model and the multifactorial model. If the methods used cannot discriminate between these extreme models, it is unlikely that they will be able to discriminate between intermediate models. Even with continuous traits there are difficulties in deciding whether a major gene is involved or in estimating the number of loci that are influencing the character (Elston and Stewart, 1973; MacLean *et al.*, 1974). These difficulties are bound to be greater with 0, 1 characters such as diseases that can be treated as having a single level of severity. Of course, it will never be possible, without identifying the loci concerned, to prove that a certain genetic model applies. It may be possible to disprove certain models if they provide an unsatisfactory fit to the observed data. Unfortunately, in practice the observed data may be biased from various factors, such as familial environmental effects, genetical and clinical heterogeneity, errors in diagnosis and in parentage and biases in ascertainment of families and in estimation of population incidence. The discovery and use of associated variables closely correlated with liability (see Section 5.5) might make discrimination between different models easier.

The strict Mendelian one-locus two-allele model can be generalized as follows:

Genotypes	A_1A_1	A_1A_2	A_2A_2
Frequency	q^2	$2q(1-q)$	$(1-q)^2$
Proportion manifesting the disease	f_{11}	f_{12}	f_{22}

A recessive gene with complete penetrance would have $f_{11} = f_{12} = 0$, $f_{22} = 1$ and a dominant gene with complete penetrance $f_{11} = 0$, $f_{12} = f_{22} = 1$. The expected proportions of each genotype among affected individuals and the expected frequency of the genotypes (and so of the disease) in relatives of affected individuals can be derived (Campbell and Elston, 1971; James, 1971). Note that this model assumes that there are no other genetic factors, common to relatives, to modify the expression of the major locus. This is biologically unlikely. It has been shown with laboratory animals that penetrance can often be modified by selection. However, it provides an extreme model with which to contrast the multifactorial model.

Elston and Campbell (1970), Wilson (1971) and Kidd and Cavalli-Sforza (1973) have fitted the above model, using maximum likelihood methods, to familial frequency data on schizophrenia and obtained a reasonable fit with several parameter sets (for example, $q = 0.06$, $f_{11} = 0$, $f_{12} = 0.08$, $f_{22} = 1.0$, i.e. all homozygotes and 8 per cent of heterozygotes for a particular allele manifest the disease). Chung *et al.* (1974) have also applied the model to sibship segregation data on cleft lip (with or without cleft

palate) with results of a similar form. The multifactorial model with only two para-
meters also gave a reasonable fit to the data, so no resolution between the models was
possible. However, the best fitting multifactorial models did imply higher recurrence
risks in sibships than the single-locus model.

Krüger (1973) and Smith (1971b) have tried to determine in what situations dis-
crimination between these extreme models of inheritance would be possible. Their
approach was to generate data by computer on one model and to test the fit achieved
to it by the other model, and vice versa. It was found that even with large numbers
of individuals and neglecting sampling fluctuations, for a wide range of situations one
model could generally satisfactorily fit (as judged by χ^2 goodness of fit tests) the data
generated by the other model. This was true both for the data on the incidence in
particular relatives aggregated over families, as well as for data on segregation within
sibships. The single-locus model could always fit data from the multifactorial model,
except when, as for many of the "common" congenital abnormalities in man, the
correlation between relatives was high and the incidence of the disease was low. On
the other hand, if a fairly strict Mendelian situation applied, the multifactorial model
would usually be readily rejected. In summary, a set of parameters for the single-
locus model could usually be found which fitted the multifactorial data, but not vice
versa. Some of the parameter sets obtained for the single-locus model were rather
extreme, and not very acceptable biologically. It may therefore be appropriate to
include a term for the prior probability of obtaining an extreme parameter set and
weight the likelihood of the two models appropriately.

Some theoretical insight into the problem of model discrimination was given by
James (1971), who showed that there were only three independent estimable para-
meters in the generalized single-locus model, namely the disease incidence and the
additive and dominance components of the genetic variance in incidence. Thus an
infinite set of $q, f_{11}, f_{12}, f_{22}$ values will fit any set of data that can be fitted at all. The
single-locus model cannot give rise to any epistatic genetic variance (i.e. variance due
to interaction between loci) while the multifactorial model can. However, with the
multifactorial model epistatic variance in incidence is only large when the heritability
is high and the disease incidence is very low (Dempster and Lerner, 1950) and so
discrimination may only be possible in these situations, confirming the results from
the computer simulations described above.

A method to discriminate between the models, depending on identifying different
levels of manifestations of the disease, or equivalently multiple thresholds, has been
proposed by Reich *et al.* (1972). With two or more thresholds the expected familial
frequencies, especially for MZ twins, may be different for the two models and so dis-
crimination may be possible. They applied their model to data on the number of
induced lung tumours in mice. Selecting two threshold levels (11 and 18 tumours,
respectively) they were able to reject ($P < 0.005$) a single-locus model but not the multi-
factorial model of liability. In practice, it may be difficult to identify two thresholds,
or to get them far enough apart, for adequate discrimination or to ensure that the
thresholds occur on the same scale of liability (see Section 5.4). Moreover, rejection of
the single-locus, two-allele model does not exclude other simple models such as one
locus with multiple alleles, or two loci.

So as to use concurrently all the information in each family, Elston and Stewart
(1971) have developed a generalized method of analysing family histories. They have
given methods for writing the likelihood of a pedigree and then of finding the maximum
likelihood estimators of parameters given data from a number of families. Different

modes of inheritance can be proposed and the fit to different models compared. By using all the information in the pedigree concurrently these methods should be more powerful in discriminating between different modes of inheritance than any of the methods described previously.

Another approach has been developed by Morton and MacLean (1974). Their concern is to test if, in addition to many loci with small effects, a locus with major effects exists. So they have specified a model with both multifactorial and single-locus components. Rather than rely only on the binomial variate, normal or affected, they have included a continuous trait, such as glucose tolerance levels in diabetes, in defining the liability of an individual. Analysis of a continuous variate may be five to ten times more powerful in detecting a major locus, than for a binomial trait, and initial trials with simulated data suggest that any loci with major effects, defined as more than 1 standard deviation difference between the two homozygotes, could be detected (MacLean *et al.*, 1974).

These and other methods are being used to try and resolve the inheritance of many familial diseases. However, the biological interpretation of the statistical results may not be satisfactory or reliable. Only when individual loci associated with the disease can be identified will the question be resolved. This will depend more on laboratory and clinical research, rather than statistical analysis. A good example of this is given by the condition ankylosing spondylitis. This disease has a frequency of 4 per 1,000 with recurrence risks in first-degree relatives of about 4 per 100. Previous genetic analyses had concluded that the condition was inherited either as an autosomal dominant with penetrance of 70–80 per cent or as a multifactorial disorder with a heritability of liability of $70 \pm 9 \cdot 3$ per cent (Emery and Lawrence, 1967). However, recently Brewerton *et al.* (1973) found a very close association of the disease with the allele W27 at the histocompatibility locus HL-A, with 90–95 per cent of affected individuals having this allele compared with about 5 per cent in the normal population. Thus the disease is largely inherited as an autosomal dominant with a penetrance of 8 per cent in males and 1 per cent in females. This major locus affect was not detected by the methods described earlier. Actually, frequencies in second- and third-degree relatives give heritability estimates over 100 per cent, but the standard errors are large and biases in these relatives are difficult to discount. This example highlights the danger in concluding that inheritance is multifactorial simply because reasonable heritability estimates are obtained.

In the current state of knowledge about many diseases, the choice that may have to be made is between a single-locus model with incomplete penetrance and a multifactorial model. The multifactorial model with its two parameters, μ and σ, or P_1 and ρ^* (h^2), often appears to be adequate to describe the available facts—the population frequency, the risk to first-degree relatives and other relations, and the concordance rates in monozygotic twins. A single-locus model with incomplete penetrance could also nearly always be found to explain the data because such a model, even with additivity at the locus, has three parameters—the risks for the two homozygotes and the gene frequency. The advantage of the multifactorial model is that it requires fewer parameters.

8. DIFFERENT GENETIC MODELS AND THEIR EFFECTS ON ESTIMATING RECURRENCE RISKS

The recurrence risks for the single locus model can be derived from first principles from the Mendelian frequencies and segregation ratios (e.g. Elandt-Johnson, 1971,

Chapter 7). To carry out the calculations, Heuch and Li (1972) have developed a computer program, PEDIG, that gives the risks with the single-locus model for any family history. Computer programs to calculate recurrence risks with the multifactorial model have already been referred to in Section 5.1.

For simple family histories all the models will tend to give good, and therefore similar, risk estimates. This is because their parameters are directly derived from the empiric risks. The difficulties mentioned earlier in discriminating between different models do imply a certain robustness in derived risk estimates to the particular model used. Two questions arise, (1) do the different models lead to different risk estimates in more complex family histories? and (2) how well do the estimated risks compare with any empiric risks available for complex families?

Comparisons with empiric risks for complex family histories are few because of the difficulty in collecting sufficient families with several affected individuals. Multifactorial risk estimates for diabetes were in reasonable agreement with empiric risks calculated from familial data (Darlow, 1972). On the other hand, empiric risks with two or more affected relatives for cleft lip ± cleft palate (Woolf, 1971) were substantially higher than the multifactorial or single-locus risks estimates (Chung *et al.*, 1974).

Morton (1969) has suggested a further model to obtain recurrence risks for sibs when the disease risks are variable between families. Using a method due to Skellam (1948), he assumed that the distribution of risk p, between families was a Beta distribution with parameters α and β:

$$f(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\beta(\alpha, \beta)}, \quad 0 < p < 1.$$

The population incidence is $\alpha/(\alpha+\beta)$, and the frequency in sibs of an affected individual, is $(\alpha+1)/(\alpha+\beta+1)$. α and β can be calculated given values of the population incidence and the incidence in sibs of affected individuals. In general, the recurrence risk, given s sibs with r affected is $(\alpha+r)/(\alpha+\beta+s)$. The values of α and β for sibships with 0, 1 and 2 affected parents can similarly be derived. Smith (1971a) and Mendell and Elston (1974) compared the predictions of this model with the multifactorial model and found good agreement when only one or two sibs were affected but less good agreement if there were more affected individuals in the family when the multifactorial model usually gave the higher recurrence risk estimates. Van Regermorter and Smith (1974) studied risks derived by the single-locus and Beta models for a range of parameter values which were compatible with selected heritability values for disease liability. When the penetrance of homozygotes was high, the recurrence risks were fairly similar for all three models. However, as the penetrance fell the risks for the single-locus model reached a plateau equal to the penetrance level and did not increase with further affected relatives. Thus different single-locus parameter sets, fitted to the same observed data, may lead to quite different estimates of recurrence risk.

9. DISCUSSION

The main value of the multifactorial model has been as a statistical tool to summarize data on frequencies of familial disease into standard and interpretable statistics. The results for diseases can be couched in the same form, correlations or heritabilities, as for continuous traits and so are easily understood. Tests can be made between results from different relatives of those affected and from different populations. If estimates are similar they can be combined to give a single estimate of the relative

importance of heredity in the aetiology of the disease. If they differ, they may give guidance as to the mode of inheritance or to the other factors, environmental or genetic, affecting liability to the disease. An important property of the estimate of the heritability of liability is that it is not a direct function of the level of the incidence of the disease. Many of the earlier summarizing statistics, such as Penrose's K ratio, (P_2/P_1^2) in our notation, Penrose (1953), Edwards' empirical result $P_2 = P_1^{\frac{3}{2}}$, Edwards (1963) and Holzinger's (1929) "index of heritability", were not successful in separating the concepts of "heritability" and the level of disease incidence.

The multifactorial model generally gives very similar results to the single locus model with incomplete penetrance. The multifactorial approach is no more difficult computationally and is, we believe, often more plausible.

The main antagonism to the model has arisen because many workers have concluded, on finding that the model fits, that in fact inheritance must be multifactorial. This non-logical step is hard to avoid, but can lead to serious errors in interpretation. For example, several quantitative traits known to be largely controlled by a single locus with multiple alleles may mimic multifactorial inheritance (Eze *et al.*, 1974) and finding a locus closely associated with ankylosing spondylitis (see Section 7) should warn of the dangers of false inferences.

Two important uses of the model, in estimating recurrence risks and in discriminating between modes of inheritance, have been covered in the previous sections. These applications depend more on computing than on advanced mathematics. Providing a model with which to contrast simple Mendelian models of inheritance may prove important, since if tests fail to detect major loci segregating for the disease, some form of inheritance which will tend to the multifactorial model may be assumed by default, and its results and implications will apply at least as an approximation.

There is often confusion about the interpretation of heritability and the possible effects of environmental change. With a fixed relation of risk to liability, heritability here is concerned with the genetic variation in liability about the current mean and tells little about shifts in the mean value due to "environmental" changes. Falconer (1965) has suggested that the only guidance it can give is that when the heritability is high it shows that current variable environmental factors, including nutrition and forms of preventive treatment, have little effect on liability so that it may be wise to look for new and novel factors to change the mean liability and hence the frequency or severity of the disease. A criticism of the interpretation of heritability is made by Kidd and Cavalli-Sforza (1973). They showed that the choice of the underlying genetic model has an important effect on the conclusion about the "proportion of the variation in liability due to genetic factors". By fitting a single-locus, two-allele model to data on schizophrenia they found that only 10–15 per cent of the variation in liability was due to differences in mean liability between genotypes. This compared with estimates of 80 per cent or higher for the heritability of liability using the multifactorial model on the same data. suggesting a much greater importance of genetic effects. However, Kidd and Cavalli-Sforza's model is biologically unlikely since it allows only environmental factors and no genetic factors to modify the expression of the major locus. Moreover, since the genetic values of their genotypes differed substantially, their result is not due to the absence of important genetic effects, but rather to the low frequency of the deleterious allele and of the abnormal homozygote compared with the other genotypes. The operational usefulness of components of variance or of heritability, which is a ratio of variance components, calculated on the scale of incidence is far from clear. The practical relevance of heritability depends on its use

as a regression or correlation coefficient in predicting the consequences of selection applied to a population. This in turn depends on the characteristic concerned being a continuous Normally distributed variable.

The multifactorial model has been used so far largely to summarize data and its possible use in splitting a disease into various clinical sub-groups with different aetiologies (see Section 5.4) has not been fully exploited.

In many respects the multifactorial model is a simplistic and "lumping" approach and Nature is likely to be much more complex and heterogeneous. With intensifying biochemical, serological and clinical research, separate entities in familial disease are continually being identified and isolated. For example, in coronary heart disease identification of various lipo-protein fractions allow new approaches to study the inherited and environmental factors associated with the disease. Thus the role of the multifactorial model in familial disease may be as a temporary tool useful during a period of ignorance for estimating risks and for providing indicators about the relations between different diseases and the relation of diseases with measurable continuous characters. Major breakthroughs must come from more fundamental research. What are the familialy correlated elements of liability and what are the familialy independent components that determine the incidence or non-incidence of disease at a given level of liability? These will be the important questions in the future.

REFERENCES

AITKEN, A. C. (1934). Notes on selection from a multivariate normal population. *Proc. Edinb. Math. Soc.*, **4**, 106.

BONAITI-PELLIÉ, C. and SMITH, C. (1974). Risk tables for genetic counselling in some common congenital malformations. *J. Med. Genet.*, **11**, 374.

BREWERTON, D. A., CAFFREY, M., HART, F. D., JAMES, D. C. O., NICKOLLS, A. and STURROCK, R. D. (1973). Ankylosing spondylitis and HL-A27. *Lancet*, **1**, 904.

CAMPBELL, M. A. and ELSTON, R. C. (1971). Relatives of probands: models for preliminary genetic analysis. *Ann. Hum. Genet.*, **35**, 225.

CARTER, C. O. (1969). Genetics of common disorders. *Brit. Med. Bull.*, **25**, 52.

CARTER, C. O., DAVID, P. A. and LAURENCE, K. M. (1968). A family study of major central nervous system malformations. *J. Med. Genet.*, **5**, 81.

CARTER, C. O. and EVANS, K. A. (1969). Inheritance of congenital pyloric stenosis. *J. Med. Genet.*, **6**, 233.

—— (1973). Spina bifida and anencephalus in Greater London. *J. Med. Genet.*, **10**, 209.

CAVALLI-SFORZA, L. L. and BODMER, W. F. (1971). *The Genetics of Human Populations*. San Francisco: Freeman.

CHING, G. H. S., CHUNG, C. S. and NEMECHEK, R. W. (1969). Genetic and epidemiological studies of clubfoot in Hawaii. Ascertainment and incidence. *Amer. J. Hum. Genet.*, **21**, 556.

CHUNG, C. S., CHING, G. H. S. and MORTON, N. E. (1974). A genetic study of cleft lip and palate in Hawaii. II. Complex segregation analysis and genetic risks. *Amer. J. Hum. Genet.*, **26**, 177.

CRITTENDEN, L. B. (1961). An interpretation of familial aggregation based on multiple genetic and environmental factors. *Ann. N.Y. Acad. Sci.*, **91**, 769.

CURNOW, R. N. (1964). The effect of continued selection of phenotypic intermediates on gene frequency. *Genet. Res.*, **5**, 341.

—— (1972). The multifactorial model for the inheritance of liability to disease and its implications for relatives at risk. *Biometrics*, **28**, 931.

CURNOW, R. N. (1974). The use of additional information in calculating disease risks from family histories. *Biometrics*, 30, 655–665.

CURNOW, R. N. and DUNNETT, C. W. (1962). The numerical evaluation of certain multivariate normal integrals. *Ann. Math. Stat.*, 33, 571.

DARLOW, J. (1972). Estimation of empiric risks of diabetes mellitus. Unpublished Hons. Thesis, Department of Human Genetics, Edinburgh University.

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

DRAPER, G. J. (1974). Some extensions to Weinberg's general proband method for estimating familial risks of disease (submitted).

EDWARDS, J. H. (1960). The simulation of Mendelism. *Acta. Genet. Stat. Med. (Basel)*, 10, 63.

—— (1963). The genetic basis of common disease. *Amer. J. Med.*, 34, 627.

—— (1969). Familial predisposition in man. *Brit. Med. Bull.*, 25, 58.

ELANDT-JOHNSON, R. C. (1971). *Probability Models and Statistical Methods in Genetics*. New York: Wiley.

ELSTON, R. C. and CAMPBELL, M. A. (1970). Schizophrenia: evidence for the major gene hypothesis. *Behav. Genet.*, 1, 3.

ELSTON, R. C. and STEWART, J. (1971). A general model for the genetic analysis of pedigree data. *Hum. Hered.*, 21, 523.

—— (1973). The analysis of quantitative traits for simple genetic models from parental, F_2 and backcross data. *Genetics*, 73, 695.

ELSTON, R. C. and YELVERTON, K. C. (1975). General models for segregation analysis. *Amer. J. Hum. Genet.*, 27, 31–45.

EMERY, A. E. H. and LAWRENCE, J. S. (1967). Genetics of ankylosing spondylitis. *J. Med. Genet.*, 4, 239.

EZE, L. C., TWEEDIE, M. C. K., BULLEN, M. F., WREN, P. J. J. and EVANS, D. A. P. (1974). Quantitative genetics of human red cell acid phosphatase. *Ann. Hum. Genet.*, 37, 333.

FALCONER, D. S. (1965). The inheritance of liability to certain diseases estimated from the incidence in relatives. *Ann. Hum. Genet.*, 29, 51.

—— (1967). The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes. *Ann. Hum. Genet.*, 31, 1.

GOTTESMAN, I. I. and SHIELDS, J. (1967). A polygenic theory of schizophrenia. *Proc. Nat. Acad. Sci.*, 58, 199.

GRUNEBERG, H. (1952). Genetical studies on the skeleton of the mouse. IV. Quasi-continuous variations. *J. Genet.*, 51, 95.

HESTON, L. L. (1966). Psychiatric disorders in foster home reared children of schizophrenic mothers. *Brit. J. Psychiat.*, 112, 819.

HEUCH, I. and LI, F. H. F. (1972). PEDIG—a computer program for calculation of genotype probabilities using phenotype information. *Clin. Gen.*, 3, 501.

HOLZINGER, K. T. (1929). The relative effect of nature and nurture on twin differences. *T. Educ. Psychol.*, 20, 241.

IDELBERGER, K. (1939). Die Ergebnisse der Zwilling forschung beim angeboren Klumpfuse. *Verh. dt. orthop. Ges.*, 25, 272.

JAMES, J. W. (1971). Frequency in relatives for all-or-none trait. *Ann. Hum. Genet.*, 35, 47.

KIDD, K. K. and CAVALLI-SFORZA, L. L. (1973). An analysis of the genetics of schizophrenia. *Soc. Biol.*, 20, 254.

KRÜGER, J. (1973). Discrimination between multifactorial inheritance with threshold effect and two-allele single-locus hypothesis. *Humangenetik*, 17, 181.

McKUSICK, V. A. (1971). *Mendelian Inheritance in Man*, 3rd ed. London: Heineman.

MacLEAN, C. J., MORTON, N. E. and LEW, R. (1974). Analysis of family resemblance. IV. Operational characteristics of segregation analysis. *Amer. J. Hum. Genet.* (submitted).

MENDELL, N. R. and ELSTON, R. C. (1974). Multifactorial qualitative traits: genetic analysis and prediction of recurrence risks. *Biometrics*, 30, 41.

MORTON, N. E. (1969). Segregation analysis. In *Computer Applications in Genetics* (N. E. Morton, ed.). Honolulu: University of Hawaii Press.

MORTON, N. E. and MacLEAN, C. J. (1974). Analysis of family resemblance. III. Complex segregation of quantitative traits. *Amer. J. Hum. Genet.* 26, 489.

MORTON, N. E., YEE, S., ELSTON, R. C. and LEW R. (1970). Discontinuity and quasi-continuity: alternative hypotheses of multifactorial inheritance. *Clin. Genet.*, 1, 81.

MORTON, N. E., YEE, S. and LEW, R. (1971). Complex segregation analysis. *Amer. J. Hum. Genet.*, 23, 602.

NEWCOMBE, H. B. (1964). (Remarks included in panel discussion of the session epidemiological studies.) In *Papers and Discussions of the Second Internation Conference on Congenital Malformations* (M. Fishbein, ed.). New York: International Medical Congress.

PEARSON, K. (1900). Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. *Phil. Trans. Roy. Soc. A*, 195, 79.

—— (1904). The inheritance of the mental and moral characters in man. *Biometrika*, 3, 131.

—— (1914). *Tables for Statisticians and Biometricians.* Cambridge University Press.

PENROSE, L. S. (1953). The genetical basis of common disease. *Act. Genet. Stat. Med. (Basel)*, 4, 257.

RAO, D. C. and MORTON, N. E. (1973). Large deviations in the distribution of rare genes. *Amer. J. Hum. Genet.*, 25, 594.

VAN REGERMORTER, N. and SMITH, C. (1974). The importance of determining the mode of inheritance for the estimation of recurrence risks (to appear in *J. Gen. Hum.*)

REICH, T., JAMES, J. W. and MORRIS, C. A. (1972). The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. *Ann. Hum. Genet.*, 36, 163.

ROBERTSON, A. (1956). The effect of selection against extreme deviants based on deviation or on homozygosis. *J. Genet.*, 54, 236.

ROBERTSON, A. and LERNER, I. M. (1949). The heritability of all-or-none traits: viability of poultry. *Genetics*, 34, 395.

ROSENTHAL, D. (1970). *Genetic Theory and Abnormal Behavior*, p. 127. New York: McGraw-Hill.

SCHULSINGER, F. (1972). Psychopathy; heredity and environment. *Int. J. Mental Hlth*, 1, 190.

SKELLAM, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *J. R. Statist. Soc. B*, 10, 257.

SMITH, C. (1970). Heritability of liability and concordance in monozygous twins. *Ann. Hum. Genet.*, 34, 85.

—— (1971a). Recurrence risks with multifactorial inheritance. *Amer. J. Hum. Genet.*, 23, 578.

—— (1971b). Discrimination between different modes of inheritance in genetic disease. *Clin. Gen.*, 2, 303.

—— (1972). Computer programme to estimate recurrence risks for multifactorial familial disease. *Brit. Med. J.*, 1, 495.

—— (1974). Concordance in twins: methods and interpretation. *Amer. J. Hum. Genet.*, 26, 454.

SMITH, C., FALCONER, D. S. and DUNCAN, L. J. P. (1972). A statistical and genetical study of diabetes. II. Heritability of liability. *Ann. Hum. Genet.*, 35, 281.

SMITH, C. and MENDELL, N. R. (1974). Recurrence risks from family history and metric traits. *Ann. Hum. Genet.*, 37, 275.

THOMPSON, R. (1972). The maximum likelihood approach to the estimate of liability. *Ann. Hum. Genet.* 36, 221.

TRIMBLE, B. V. (1971). An empirical simulation of quasi-continuous inheritance. In *Excerpta Medica.* Inter. Congress Series No. 233, 4th Inter. Congress of Hum. Genetics, Paris, 1971.

WILSON, S. R. (1971). Fitting of models of incomplete penetrance to family data. *Ann. Hum. Gen.*, 35, 99.

—— (1974). Fitting of multifactorial models to family data. *Ann. Hum. Genet.* 38, 231.

WOOLF, C. M. (1971). Congenital cleft-lip. A genetic study of 496 propositi. *J. Med. Genet.*, 8, 65.

WORLD HEALTH ORGANISATION (1972). Report of a Scientific Group. Genetic disorders: prevention, treatment and rehabilitation. Tech. Rep. 497, WHO, Geneva.

WRIGHT, S. (1934). An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genet.*, 19, 506.

WYNNE-DAVIES, R. (1970). The genetics of some common congenital abnormalities. In *Modern Trends in Human Genetics* (A. E. H. Emery, ed.), Vol. 1. London: Butterworth.

DISCUSSION OF THE PAPER BY PROFESSOR CURNOW AND DR SMITH

Dr C. O. CARTER (M.R.C. Clinical Genetics Unit): I am very happy to propose a vote of thanks to Professor Curnow and Dr Smith. My background to this is almost entirely non-mathematical, but I have been collecting family data for years on common congenital malformations and other reasonably common conditions which clearly have some degree of genetic determination. These curious features, which the authors have described, were early apparent: first, the fact that consistent family patterns were obtained which were

different from those from straightforward Mendelian inheritance; secondly, that the relatives of the less commonly affected sex were more often affected than those of the more commonly affected sex; thirdly, the cumulative effect, in that if two individuals in a family were affected already there was a higher recurrence risk than when one member only was affected; fourthly, the effect of the severity of the malformation on the recurrence risk—and so on.

Early on, in the 1960s, thinking about these points in relation to pyloric stenosis and cleft lip, I proposed a multifactorial aetiology with the genetic component being polygenic. But I do not have the mathematical ability to work this out properly. In 1960 John Edwards had already made a major contribution, which I had not read at the time or I would have received some help from it.

I was delighted when the agricultural geneticists, who were much more familiar with this kind of subject than we were in relatively simple medical genetics, moved into the field. Professor Falconer's first paper was a revelation to me, explaining and quantifying many of the observations I had made. It gives me great pleasure to see this being developed further.

I think some difficulties remain—again, I was pleased to see the discussion of twin concordance in tonight's paper. I felt instinctively that a high twin concordance was not needed on any threshold model because the pairs could be nicely balanced on either side of the threshold, but it was interesting to see it quantified. However, it is still something of a problem, particularly with congenital malformation, because not only do monozygotic twins have the same genotype, but they *prima facie* grow up in precisely the same intra-uterine environment—yet there is still only 20–30 per cent concordance for many of these common malformations. It might be expected to be higher because of the common environment.

For genetic counselling, clearly we look with interest at the predictions based on developments of the multifactorial model. Again, I think we must be a little cautious here because, as Professor Curnow and Dr Smith said, these conditions may be heterogeneous and nearly always there is a relatively rare component of single gene conditions mixed up with the multifactorial which cannot always be distinguished. We are able increasingly to distinguish them; for example, there is a type of cleft lip and cleft palate which behaves as the simple dominant condition, which can be picked out because there are mucous pits in the lower lip. When we see the kind of family in which the father has had cleft lip, also two of his children and, later, one of these children has had an affected child, even though there is no mucous pit present we have to ask ourselves whether we are dealing with yet another single gene condition, one which cannot be distinguished. We would tend to give a higher risk to a relative than that calculated from the polygenic, multifactorial model.

It seems to me that tonight's authors may not have brought out sufficiently the value of the second- and third-degree relatives in distinguishing between the multifactorial and the modified single gene inheritance. It is in the second- and third-degree relatives that the heritabilities of over 100 per cent begin to appear if the polygenic model is applied to a modified single gene condition. For some conditions—not all—we have good data on second- and third-degree relatives; for example, cleft lip.

Overall, however, looking at these mathematical developments from the sidelines I have found the last 12 or 13 years extremely stimulating and enjoyable. I should like to thank our two contributors tonight for giving such an excellent review of the developments over this period of time and have great pleasure in proposing the vote of thanks.

Professor J. H. EDWARDS (The United Birmingham Hospitals): It is with some hesitation that I accepted this challenge. I am a consumer, rather than a producer, and, since I am involved in advising patients with familial disorders, I would welcome any numerical aid which would lighten my burdens and clarify theirs. I feel in the position of having to

decide whether we are dealing with something which, to take two broad categories in a similar situation, may be chess or warfare. Are we dealing with a highly formalized situation, which has to be accepted and discussed for its elegance and intellectual stimulation, or with something crude and practical? Not being able to follow some of the deeper incursions into the infinitesimal calculus, I find myself in the position of an experienced general and amateur chess player who is asked to comment on a game between experts.

I should like to take the empirical view, which seems to be a tradition of this Society, and ask the question whether this approach bears fruit or flowers. To my mind, the authors have given us an elegant procedure which bears flowers, but I do not feel that it bears fruit. As chess is regarded as a more intellectual activity than warfare, and flowers more elegant than fruit, this need be no criticism.

There are three points on the "fruit" which I shall discuss.

First, I do not regard it as a matter of fact that there is an increasing proportion of familial disease. This has always struck me as a surprising statement because the diseases of the past, which were mainly of infectivity and of unhealthy and hostile environments, are intensely familial. The Brontës coughed all over each other: Edward Gibbon's six succeeding brothers—all, for good measure, called Edward Gibbon— died in infancy. The diseases are becoming mellower now, and they are probably becoming less familial.

I think that the concept of familial disease is a confusing one because it is difficult to imagine a non-familial disease. We are all exposed to the conditions to which the flesh is heir, and there are, by definition, no disorders in man—or likely to be—to which the flesh is not heir. I cannot conceive of a disease irrelevant to the genetic background.

Secondly, the question of models. There is the implication that some statisticians consider that biologists are wandering about trying to fit models to things. In fact, in the days of the ultracentrifuge, the electron microscope, and the genetic code, the time for this is past, just as it is past in such subjects as geography, mechanics and the planetary system. We have a basic model, and the residual problems are of estimation and not of decision. We are not really in the position of a blind man wandering about looking for shoes, who goes into a hat shop and finds they do not fit, and then into a glove shop with the same result, and eventually goes into a shoe shop and feels that he has arrived. We are in the position of somebody who goes straight into a shoe shop, and then has problems of exact fit. There is no problem of genetic models: we have a complete Lego set provided for us by the molecular biologists, which leads to the expected consequences. We have one set of Lego only—there are no alternative models; all these so-called models grade into each other and provide only estimation problems.

One way in which we can try to plot this is rather simple (Fig. 1). If we have a large number of loci, we can take the effect of any allele at any locus, as a_t, and frequency at that locus as p_t. This can be summed over each locus, and also over all loci. The loci
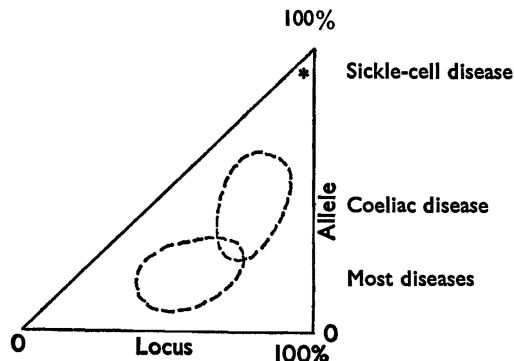

FIG. 1.

are very large in number, running to hundreds of thousands—there is room for millions. The alleles are more limited in number, so that, at most loci, any pair chosen at random are likely to be identical.

Given that this is our only endowment, we have to ask ourselves what is the strongest allele (which a_i is the greatest—where a_i is the ith allele), which is the most influential allele at one locus (which $a_i p_i$ is greatest) and which is the most influential locus, at which is the sum of $a_i^2 p_i$ greatest. These are all slightly different, but are questions which need to be asked.

The specification of "amount" is difficult. One solution which is simple, although it has not been very productive, is to use contributions to variance. Once this specification is made we can plot the proportion due to the strongest or the most influential allele and also to the most important locus, both going from zero to 100 per cent.

If we take a disease which is inevitably the consequence of some specific allele, for instance, sickle-cell disease or Tay–Sachs disease, then it is represented by a point at the apex of the triangle as shown.

There are some diseases in which there are many alleles at one locus which considerably influence the incidence; for instance, coeliac disease and myasthenia gravis, which are greatly influenced by alleles at, or near, the HL-A loci. There are many alleles and the strongest must be somewhere near the point shown. All the diseases to which the flesh is heir must lie within this triangle; all of them must have a strongest allele, and a strongest locus. It seems that we have an estimation problem at this level, not a decisional one. The problem is to find where they are.

This makes a difficult and very interesting estimation problem—and estimation appears to be what is needed in advanced sciences, in which there are few remaining yes–no questions. Genetic linkage is a scientific activity in which significance tests can be justified, but there are now few fields of biology in which a significance test can be done without admitting a degree of ignorance which is inappropriate.

The third point at issue is how to give an opinion to somebody who is asking for advice about a recurrence risk. There are various genetic correlations, or relationships: there is the sib/sib, and the parent/child genetic correlation which is of 0·5, and the cousin/cousin relationship of 0·125. In an ideal situation of complete genetic determination, the phenotypic correlation will be equal to the genetic correlation. People are worried about the risks of disease, first, because they have seen the diseases from which their relatives have suffered, they know it is unpleasant and think of it as *the* disease about which they have to worry. Secondly, because they know that diseases tend to run in families. At least, most people seem to know this.

There are some extremely good data for this purpose, much of it collected by Dr Cedric Carter. Given a defined relationship to a victim of some disease, the risk of affliction necessarily is monotonic with the degree of resemblance defined by this relationship. We have this sequence of risk, which is never 100 per cent, although it may be 50 per cent or so in an identical twin. If we take the logarithm of this incidence, this is linear against the phenotypic correlation on the exponential or logistic model (Edwards, 1968) (Fig. 2). So we can plot a line through a series of points connecting the extremes of unrelated individuals (the population incidence) and the incidence in identical twins. It does not have much meaning, and it is difficult to conceive a relationship which would give a genetic correlation in the intervening levels, but a regression line can be plotted. The practical point is, given the data points and the line, on which is the advice to be based. I think that the advice should always be given on the data, not on the regression line, so I do not find this concept of regression of practical utility. Fortunately, the level is so low that it does not usually cause problems. But this is a situation in which one has to work on the raw data, and there is no way in which those raw data can be "cooked" and made more useful by taking, not a datum, but the intercept of the regression at that point.

A further point in relation to using the term "heritability" in this context is that it is an emotive word because it sounds rather like heredity and it is easy to obtain the impression

that they are simply connected. Of course, they are connected if one is experienced at adding together the variances but, in the mundane world in which distinguished observers such as Jensen require an armed guard—partly because they themselves fail to appreciate the formal irrelevance of heritability values to opportunities for environmental benefit—it must be accepted as easily misunderstood.
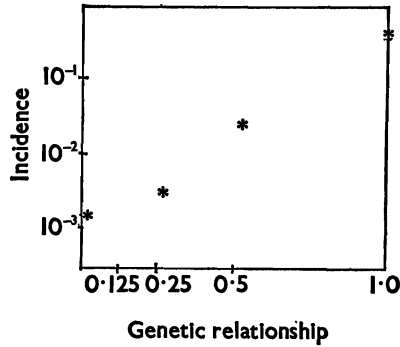

Genetic relationship
FIG. 2.

The animal geneticists use the parameters environment squared (e^2) and genetic contribution squared (g^2), and define the heritability (h^2) which is the proportion the genetic contribution bears to the sum of both. This is useful in predicting the degree of response to breeding. In man it is rather more confusing, although the alphabet is very helpful. There is a parameter, which might be called "domesticity", d^2, which is the non-genetic familial environment. There is g^2, the genetic component. These two added together give the familiarity or f^2. There is also the non-familial environment e^2: that is, the environment in the classical sense. We therefore have a large number of parameters and, in practice, usually just two degrees of freedom. (The bulk of the information in a pedigree can be summarized by the three possible types of unordered sib-pair.) What can be estimated in man is not heritability, but familiarity. This has a number of advantages as a concept because, even if 100 per cent, it does not imply any limitations on environmental response.

Having said that I did not feel that the fruit was actually fruitful, I might go on to an aspect which I find more exciting and interesting, perhaps because I am less able to understand it: that is, to the flowers. We have to try to provide some simple way of handling these very complicated functions, and the Gaussian curve itself is quite complicated. Indeed no two statistical books seem to give exactly the same method of handling it, in part, because it is written in a peculiar form. It can be written

$$1/\sqrt{(2\pi e^{x^2})}.$$

This makes it a little easier to understand but, even so, it is complicated and the usual method of integration is confusing, requiring acts of faith even in advanced textbooks. Obviously, if there are two variates it is even more complicated and, in the general case, has no algebraic solution. It is surprising that this problem has become such an unfashionable pursuit because, in 1897, Sheppard made the interesting observation—which he proved—that in the symmetrical case

$$r = \cos\{a/(a+b)\}\pi.$$

This is rather remarkable as it might be thought that it would never equal one; this is no problem because in the familial case it is never zero as it becomes increasingly elliptical.

This leads to the approximation $z = (\pi/8) \log_e (ad/bc)$† which is empirically very robust, rather surprisingly, and it does not seem to matter much where the thresholds are (Edwards, 1957, 1969). This is an important point because the effect of the environment is to move the thresholds around. Thus, in the case of this identity, it would seem to me that heritability is formally irrelevant to the possibilities of environmentally improving a disease—tuberculosis is now so rare that its infection is difficult to acquire, although Karl Pearson thought that immigration should be suppressed because this highly *heritable* disease would become rampant due to the *genetic* contribution of the various groups of immigrants who were coming into the country at that time. There are excellent reasons for stopping the immigration of people with tuberculosis, but "heritability" is hardly relevant.

In 1904 Karl Perason, with his remarkable aptitude for integrals, solved the bivariate problem with the tetrachoric coefficients, and explored and tabulated this model which Professor Curnow kindly attributed to me—but this was in 1904, 65 years earlier. He and Everitt tabulated this (Everitt, 1910), and their infinitesimal integrations were confirmed by Dr Smith's electronic enumerations. It might seem like fishing with a worm to check on Karl Pearson's algebra with a computer, but those of us who want fish are very glad that this has been done.

Finally, there is the difficult problem, the three-dimensional system in which there are parents and a child, the four-dimensional system when there are two children. This defeated Karl Pearson and empirical studies are impractical on a computer because, if sufficient slices are taken, there is insufficient time and space. It now seems to have been solved completely, in special cases, although I am afraid that I cannot follow Professor Curnow fully on his algebra. This remarkable achievement seems to give results which are consistent with observations. The only disturbing feature is that some of the lines in their figure cross.

I should like to thank the authors very much for their "flowers"; to be distinguished for both flowers and fruits is too much for most plants, and is hardly to be expected in the field of mathematics. I hope that in the published edition of this paper Professor Curnow will give a little more space to his integral, since it is elegant and brief, and is part of the solution to this extremely difficult problem of the generalization of the multivariate normal hypersurface.

It is a pleasure to second a vote of thanks for a paper which so concisely describes what would be expected on plausible assumptions, and which has, in part, been confirmed by the empirical studies of integration over finite intervals.

The vote of thanks was passed by acclamation.

Mr G. J. A. STERN (London N.6.): I would hesitate to comment on this paper, had the authors not brought in some non-statistical considerations which affect us all. I mean that they appear to advocate selective abortion for cases where there is an appreciable risk of the child being born with a partly hereditary disease or malformation. We are talking about ending a life on the grounds of a possible disability, and we need to see with what degree of certainty such disability can be predicted.

The recent controversy popularly associated with the names of Jensen and Eysenck on inheritance of I.Q. level shows that there is by no means universal agreement either on the degree of heritability of some qualities or on the models and methods used to establish that degree. To the outsider, it looked as if convincing arguments could be made on both sides, at least as far as the racial aspects were concerned: would this be the case here if similar political passions were aroused?

The authors give correlation in liability for some abnormalities and defects, but these correlations are usually low; often only 0·3–0·5. Some of the higher levels quoted relate

† Where the symbols a, b, c, d, refer to the volumes of a doubly dichotomized bivariate distribution.

to twins, which, while very relevant to the theory, have no direct applicability to genetic counselling, for one cannot selectively abort one twin because the other has a defect. The authors admit that they cannot verify which of several models they use is the most appropriate, and inside the models they detail quite far-reaching assumptions as to distributions etc. They place weight on several studies made on the heritability of schizophrenia. Yet schizophrenia is difficult to diagnose with certainty, and one can imagine that in many cases knowledge of the fact that the parent was schizophrenic may have affected the diagnosis. In short, it seems that in most cases all that can be said is that there is an appreciable probability of the child being born with a disability, but that the odds in favour of a healthy child are still far better than even. Is there not something repulsive and all too reminiscent of Nazi ideology and practice in abortion on such grounds? What will surviving children feel about parents who would have had them aborted on suspicion in this way?

It seems to me that such studies are valuable (not that my praise or blame is of consequence), but that there is danger of forgetting the human dimension. The seriousness of the condition is hard to predict, as is the degree to which the person, or advances in medical science, can overcome it. Let us recall the case of Christy Brown, the celebrated Irish novelist. I daresay that not one case in 100 of those cited by the authors was as hopeless as Christy Brown's—paralyzed except for one foot, and unable to talk comprehensibly, so that he was long thought to be mentally retarded. Yet he is now a best-selling novelist (operating an electric typewriter with his foot) and has got married. He is fortunate in having escaped the genetic counsellors, and society is fortunate that he did so.

It is also true that quite apart from medical advances, the human being can often overcome many serious conditions so that hardly any trace remains. Yet all these possibilities for improvement, at least as significant and beneficial as the Moon expeditions, will be literally aborted if current policies continue. Let us find out about heritability of disease, certainly, but let us not use this knowledge for the easy way out of final solutions which impoverish the human race and strangle advances in healing and care.

The following contributions were received in writing, after the meeting:

Dr O. Mayo (Waite Agricultural Research Institute, University of Adelaide): In Section 6, discussion of cystic fibrosis as a possible selectively balanced polymorphism does not take account of (a) the fact that the 2 per cent advantage required to maintain the disease at its current frequency is far less than the presumptive advantage conferred by the observed differences between genotypes in fertility (which were claimed to be the source of the heterozygous advantage), (b) the possible effects of population size on the frequencies of such traits (Robertson, 1962) and (c) the probability that there is genetical heterogeneity in this disease (e.g. Polley and Bearn, 1974). It is still fair to say that the polymorphisms associated with malaria remain the only ones where convincing evidence of balanced polymorphism exists.

The discussion of the associations between polymorphism and disease, while directed towards discrimination between different modes of inheritance (Section 7), does not do full justice to the vast body of data on such associations. That between HL-A and ankylosing spondylitis is only the most recent and dramatic, allowing, as it apparently does, the reclassification of this disease as unifactorial; it may well be that similar conclusions could be drawn for many others if the genetical resolution were as fine as is possible for the HL-A system. In addition, the use of information from associated or linked loci for the resolution of genetical heterogeneity and prediction of liability should not be discounted; as is implied in Section 5.4, this can readily be incorporated into the general framework used by the authors, yet they mention specifically only the newer technique of using information from associated continuous traits (Section 5.5). While the magnitudes of the risks to persons of different genotypes are rarely as disparate as in the HL-A–ankylosing spondylitis case, the differences are not everywhere negligible.

Professor T. REICH (Washington University): The excellent description of multi-factorial models presented in this paper suggests a number of comments concerning the analysis of family data for psychiatric disorders. In general, the multifactorial models are suitable for the analysis of these data, since at the outset they do not require the assumption that environmental effects common to relatives are absent. Furthermore, the number of parameters required to define the models is small when compared with other theoretical modes of disease transmission, allowing many hypotheses to be tested which may otherwise be approached only with difficulty.

Diagnostic validity

The first problem which must be faced in studying psychiatric disorders is the problem of diagnostic classification. For all of the major functional psychiatric disorders, several systems of classification are in use. Often heterogeneous entities are grouped as a consequence of outmoded theories about their aetiology. Usually, severe or definite cases are universally recognized, but there is poor agreement about mild or border-line cases. For example, schizophrenia is defined in Europe as a severe disorder with many persistent psychotic symptoms, and the population prevalence is approximately 0·85 per cent (Slater and Cowie, 1971). The most popular American criteria for schizophrenia include these severe cases, but also include mild cases whose illness may not be protracted. The prevalence of this wider form of "schizophrenia" is approximately 4 per cent (Kety *et al.*, 1973). Using computational techniques analogous to those described in Section 5.4 of this paper, it can be determined whether the two types of schizophrenia are drawn from the same liability distribution, or whether they represent two independent entities. The observation that these two entities can be represented along the same phenotypic dimension would validate an expanded concept of the disorder and provide additional classes of information for analysis.

Mild or subclinical analogues of major psychiatric illness, such as alcoholism, manic-depressive illness and anxiety neurosis have also been defined, and by repeated application of the anlytic techniques described in this article, an increasing proportion of the population can be defined with respect to the liability to develop these disorders. In this way, new thresholds in the liability distribution can be specified and recurrence risks can be improved, both with respect to the probability of being affected and the kind of disorder which may occur.

Since genetic components are not required, determination of the relationship between varieties of a disorder may proceed using family data where common environmental effects are present. Assumptions about the relationship between the correlations for different classes of relatives are not required (i.e. a separate correlation may be estimated for parents, siblings and half-siblings) and sources of error due to nongenetic familial effects, assortative mating and selection are minimized.

Environmental heterogeneity

The multifactorial models described here may not only be used to resolve questions of genetic heterogeneity, but also to investigate the effects of environmental heterogeneity. Using an approach suggested by Falconer (1952), the concept of a trait is broadened to include the environment in which it occurs. If a population of affected individuals is divided into two or more groups, the estimation of correlations, and cross-correlations between relatives for these "traits" may greatly improve our understanding of environmental effects on the incidence of a disorder and on its transmissibility from parent to offspring. It is possible that our understanding of what constitutes a "relevant environment" may be altered and programmes for the prevention of these disorders may be improved. It must be remembered that multifactorial models take the population prevalence into account, so that the effect of environment on the transmissibility of a disorder can be directly assessed.

Alcoholism is approximately four times as common in men as in women. Reich *et al.* (1975) used a multifactorial model to investigate the phenotypic differences between alcoholism in male and female populations and were able to conclude that nonfamilial environmental factors could entirely explain the sex-effect. By contrast, Cloninger *et al.* (1975) investigated antisocial personality in males and females and were able to show that the large sex-effect could be represented by two thresholds in the same liability distribution. These latter findings supported the view that nonfamilial environmental sources of variation were equal in the two sexes.

The additivity assumption

Assumptions of additivity made when estimating genetic components have been a persistent source of acrimonious debate in behaviour genetics. In the present context, correlations and cross-correlations between relatives can be estimated from different points on the liability distribution and the additivity assumption used in defining the liability can be tested. The detection of interactions depends on the number of available thresholds and on the distance between them. Even though the assumption of additivity is robust, major interactions may be detectable with a large sample. Heterogeneity based on severity, sex-effect and polymorphic symptomology provides natural phenotypic distinctions between affected individuals and may offer opportunities for detecting non-additive interactions. In addition, independently transmitted subvarieties of a disorder may be detected and removed from the data set, resulting in a more homogeneous residual. This approach may be contrasted with the use of carefully normalized scales for measuring behavioural traits where non-additive interactions may be concealed when the trait is normalized.

It is my opinion that the first step in understanding the genetics of a psychiatric disorder is a methodologically sound analysis of the correlations between relatives when common environmental effects are present. These analyses can be helpful in defining the entity to be studied, in recognizing independent varieties of the disorder, and in broadening the concept of a trait, so that a larger population of affected individuals can be studied. The multifactorial models of disease inheritance can be most useful in these investigations.

Professor I. I. GOTTESMAN (University of Minnesota): As a fairly satisfied user-consumer of the multifactorial models so neatly and comprehensively reviewed by Professor Curnow and Dr Smith, I would hope that their efforts would reach a deservedly large audience, larger than the Royal Statistical Society. I say this because the applicability of their methods goes beyond the $4\frac{1}{2}$ per 1,000 congenital malformations and the 20 per 1,000 infants born with significant physical malformations to many forms of mental retardation (4 and 20 per 1,000 respectively for severe and mild retardation (Roberts, 1973) and to many conditions with later ages of onset such as ulcer, diabetes, hypertension and concomitants of ageing itself (e.g. senility and variation in the age at death from so-called natural causes). I suggest supplementing their excellent list of references with a few items to emphasize these points as well as the agricultural origins in plant and animal breeding of the multifactorial models: Lerner (1958), Jinks and Fulker (1970), Mather and Jinks (1971) and Fraser Roberts (1973).

I find the given definitions of polygenic and multifactorial diseases (Section 1) perpetuating an unintended ambiguity that has haunted us since the arguments between the Mendelians and the biometricians at the beginning of the century; the definitions are not mutually exclusive. Nowadays everyone recognizes that genes have pleiotropic effects, that the phenotype of interest can be modified by environmental factors, and that the phenotype of interest can be modified by the genetic background of the gene or genes of interest. The definitional problem stems in part from the gap between population genetics and clinicians, on the one hand, and physiological genetics, on the other.

Grüneberg (1952) pointed out that "the multiple genes of quantitative genetics are in fact nothing but genes whose remote effects only are being studied" (p. 110). It is necessary to avoid confusion by noting that some phenotypic traits in man are associated with major monolocus effects, genetic background effects, and poly-environmental effects while others are associated with poly-locus (polygenic), genetic background, and poly-environmental effects; the term "multifactorial" does not help us distinguish between these two classes as acknowledged at the end of Section 2. The distinction that can be made by the locutions above is blurred at the level of gene action as well as by the recognition that some genes in a polygenic system may have much larger effects than others (cf. Wright, 1934; Thoday, 1967; Gottesman and Shields, 1972). If I have added to the confusion, I apologize.

I have reserved my main comments for uses made by Curnow and Smith of data from the study of schizophrenia in the biological and adoptive families of schizophrenic probands. I would have expected them to use data on diabetes mellitus given the greater experience with this disease and the more reliable diagnoses that provided the data (e.g. Falconer, 1967; Simpson, 1969; Smith *et al.*, 1972). Shortly after Falconer's seminal 1965 paper on the inheritance of liability to threshold diseases was published, James Shields of the Institute of Psychiatry, Maudsley Hospital, and I recognized the probable usefulness of the approach to the analysis of data we had been collecting on schizophrenic MZ and DZ twins, their co-twins and their other relatives. We were able to visit Falconer in the summer of 1966 and have worth-while discussions that led to our introducing his liability model to the behavioural sciences (Gottesman and Shields, 1967, 1968). Unresolved questions we raised were later solved by Falconer (1967) and Smith (1970). Further advances by Smith (1971a) permitted us to compare the predictions of a polygenic theory of schizophrenia with those of Slater's specific dominant gene with incomplete penetrance theory and with the appropriate pooled empirical observations from the systematically conducted studies in the literature (Gottesman and Shields, 1972). The population prevalence and the heritability of liability value we used to generate the polygenic predictions were dictated by another analysis we performed in an effort to find a convergence point from various data sets that were more extensive than those appearing in Curnow and Smith's Table 3 under our names. Fig. 3 shows the results we obtained for the heritability of the liability to schizophrenia by the triage of six different population prevalences for monozygotic and dizygotic twins, the sibs, the offspring of two schizophrenics and the second-degree relatives. As a consequence we chose a value of 1 per cent for the population prevalence and a heritability of liability of 80 per cent, the nearest tabled value in Smith (1971a), to generate predictions for the risks to probands' sibs and children as a function of the number of parents affected with schizophrenia. The results are shown in Table 1 and confirm Curnow and Smith's conclusions about the difficulty of distinguishing between single locus and multifactorial modes of inheritance for common disorders.

The points in Section 3 about the confounding effects of common familial environment are well taken but are not as well made as they might be. An expanded discussion of the problems is given by Cavalli-Sforza and Feldman (1973) and Rao and Morton (1974). The data presented in Table 2 (Section 3) do make the point that schizophrenia occurs in the biological relatives of adoptees who became schizophrenic and at much higher rates than in the adopted relatives who reared them. However, the majority of the biological relatives were half-sibs and not first-degree relatives as labeled; further, the proportion of relatives shown as affected with schizophrenia actually consists of definite plus uncertain diagnoses of schizophrenia plus a hard-to-define category of "schizophrenia spectrum disorder". The problems of dealing with the latter are formidable (cf. Shields *et al.*, 1975) even with the aid of the suggestions from Reich *et al.* (1972); cf. Cloninger *et al.*, 1975). In an update of the data on adopted Danish schizophrenics (Kety *et al.*, 1975) 173 biological relatives have been identified for the 33 probands; of the former 66 are parents, 41 are maternal half-sibs, 63 are paternal half-sibs and 3 are full sibs. The prevalences (without age-correcting) of definite and then definite plus uncertain schizophrenia in the paternal half-sibs are reported and permit further efforts at fitting to the two models in our Table 1

above. We get an embarrassment of "riches". The prevalences in half-sibs of 13 and 22 per cent correspond to 1·6 and 3·0 per cent in control half-sibs. Using Smith's (1970) graph the correlation in liability for half-sibs becomes 0·45 and then 0·56; both figures lead to heritabilities near 200 per cent. If the data are used to calculate the penetrance of a posited dominant gene (Slater and Cowie, 1971) with frequency 0·03, the values are too high to be credible. At this point of model "unfitting" we cannot tell whether the data or the models are "embarrassed".
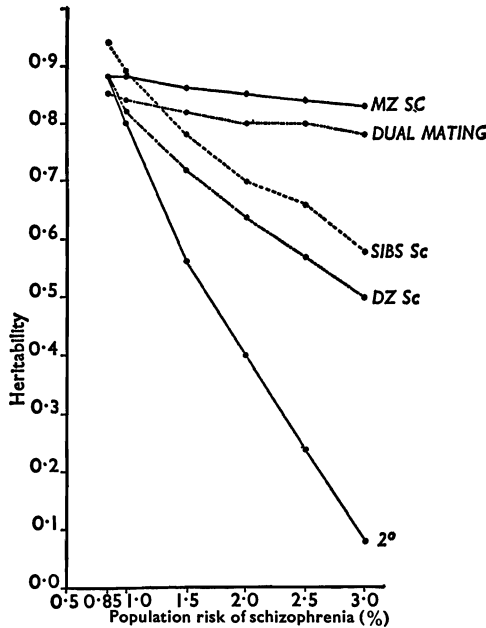

FIG. 3. Smith-type heritabilities of the liability to schizophrenia as a function of varying population risks, estimated from risks in different classes of probands' relatives (from Gottesman and Shields, 1972).

TABLE 1

Schizophrenia risk as function of parent status
(Source: Gottesman and Shields, 1972)

	Risk (%)			
	(a) *To probands' sibs*		(b) *To probands' children*	
No. of parents affected	0	1	1	2
Observed, pooled risks	9·7	17·2	13·9	46·3
Predicted, polygenic	6·5	18·5	8·3	40·9
Predicted, monogenic	9·4	13·5	8·8	37·1

Curnow and Smith provide their own best caveats about the limitations of multi-factorial model fitting when they conclude, "Thus the role of the multifactorial model in familial disease may be as a temporary tool useful during a period of ignorance for

estimating risks and for providing indicators about the relations between different diseases and the relation of diseases with measurable continuous characters. Major breakthroughs must come from more fundamental research." And earlier (Section 7), "Of course, it will never be possible, without identifying the loci concerned, to prove that a certain genetic model applies." Although it leads to unsung heroism, the methods reviewed by Curnow and Smith permit the disproof of models or at the least their thoughtful revision after further data collection. The compatibility between data and a particular model is a necessary but not sufficient condition for the credibility of the model. The heuristic value of sophisticated multifactorial models for generating collaborative research among clinicians, biostatisticians and molecular/developmental biologists must be acknowledged and praised.

The authors replied in writing as follows:

We are grateful to contributors to the discussion for their interesting comments.

We agree with Dr Carter that second- and third-degree relatives could be very useful in discriminating between different models of inheritance and have shown this empirically elsewhere (Van Regermorter and Smith, 1974). However, because of the lower degrees of relationship, large numbers of such relatives will be required, for example to get heritability estimates with low standard errors, and the information on these relatives is often less reliable than on first-degree relatives.

The value of genetic linkage and associations with genetic markers, stressed by Dr Mayo, for discriminating between simple and complex forms of inheritance is doubtful, for these have not been very productive in the past, for a summary see Mayo (1972). However, with the advent of several strong HL-A disease associations, the value of these methods may have to be reconsidered, as we indicated.

Our intentions, as applied statisticians in medical genetics, are to measure, understand and predict the occurrence and recurrence of these familial diseases. We wish to provide clinicians and families with information about possible forms of inheritance and about risks, so that individuals can make informed decisions about their families. There is neither compulsion nor advocacy, such as Mr Stern suggests, in genetic counselling, but rather, as in other fields of medicine, clinicians working to improve the health and welfare of their patients. This is not to deny the difficult moral issues associated with abortion.

Our terminology "familial disease" for a disease which "runs" in families does not please Professor Edwards but we like its generality. Common genes cannot be implicated until common familial environmental effects have been discounted. Professor Edwards states that heritability (g^2) cannot be estimated in man, but only familiarity (f^2). The whole point of including Section 3 in the paper was to consider and refute this argument and to show that (g^2) can be estimated free from (d^2), the domesticity component. Thus, if Karl Pearson had studied unrelated persons (e.g. spouses) living together, or relatives living apart, he would have found that non-genetic common familial effects were very important in tuberculosis and would have moderated his counsel accordingly. In another part of his remarks Professor Edwards suggests that because we know the basic model of inheritance (DNA and the genetic code) there can be no further models to test in genetics, and so only estimation problems remain. This would be a sorry state of affairs for any science, and would sadden Karl Popper. However, with the organizational complexity of the genome and of the phenotype deriving from it, there are many interesting models to test despite having only one form of building brick. Edwards' Fig. 1 and his remarks about HL-A illustrate this point well. Coeliac disease with a major effect of the HL-A locus now falls in the middle of his diagram. What about diabetes, schizophrenia, the congenital abnormalities and all the other familial diseases? Does a locus with a major effect exist for these conditions? This is a hypothesis we may be able to test, after defining a "major effect". The outcome of the test may then determine how to allocate genetic research effort for these diseases. Similarly we have indicated that tests for genetic identity

in related diseases, such as early and late onset diabetes, or for genetic heterogeneity within a clinical condition, may be useful.

In saying that research with the multifactorial models has borne little fruit, Professor Edwards has ignored much of the material presented and discussed in the paper. The model gives an explanation, and an understanding, of many of the empirical findings which were otherwise difficult to reconcile, and Dr Carter spoke about this in his remarks. Indeed Professor Gottesman advocates greater scope and use of the model for the wide array of familial diseases in man. Our work with estimation of recurrence risks deals, not with the simple cases indicated in Fig. 2 in Professor Edwards' remarks for which we have empiric risks, but with risks in families with a more complex family history for which there are no empiric risk estimates, and with continuous traits which may be associated with a disease. These should be useful in assessing risks of disease in relatives, so that preventive measures may be applied, and in genetic counselling about the risks to future children.

REFERENCES IN THE DISCUSSION

CAVALLI-SFORZA, L. L. and FELDMAN, M. W. (1973). Cultural versus biological inheritance: phenotypic transmission from parents to children (a theory of the effect of parental phenotypes on children's pneotypes). *Amer. J. Hum. Genet.*, **25**, 618.

CLONINGER, C. R., REICH, T. and GUZE, S. B. (1975). The multifactorial model of disease transmission: 3. familial relationship between sociopathy and hysteria (Briquet's syndrome). *Brit. J. Psychiat.* (in press).

EDWARDS, J. H. (1957). A note on the practical interpretation of 2 × 2 tables. *Brit. J. Prevent. & Soc. Med.*, **11**, (2), 73.

—— (1968). In *Record Linkage in Medicine* (E. D. Acheson, ed.) (Proceedings of the International Symposium, Oxford, July 1967). Edinburgh: Livingstone.

EVERITT, P. F. (1910). Tables of the tetrachoric functions for fourfold correlation tables. *Biometrika*, **7**, 437.

FALCONER, D. S. (1952). The problem of environment and selection. *Amer. Naturalist*, **86**, 293.

FRASER ROBERTS, J. A. (1973). *An Introduction to Medical Genetics*, 6th ed. London: Oxford University Press.

GOTTESMAN, I. I. and SHIELDS, J. (1968). In pursuit of the schizophrenic genotype. In *Progress in Human Behavior Genetics* (S. G. Vandenberg, ed.). Baltimore: Johns Hopkins Press.

GOTTESMAN, I. I. and SHIELDS, J. (1972). *Schizophrenia and Genetics—a Twin Study Vantage Point.* New York: Academic Press.

JINKS, J. L. and FULKER, D. W. (1970). A comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behaviour. *Psychol. Bull.*, **73**, 311.

KETY, S. S., ROSENTHAL, D., WENDER, P. H., SCHULSINGER, F. and JACOBSEN, B. (1975). Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic: a preliminary report based upon psychiatric interviews. In *Genetics and Psychopathology* (R. Fieve, H. Brill, and D. Rosenthal, eds). Baltimore: Johns Hopkins Press.

LERNER, I. M. (1958). *The Genetic Basis of Selection.* New York: Wiley.

MATHER, K. and JINKS, J. L. (1971). *Biometrical Genetics.* Ithaca: Cornell University Press.

MAYO, O. (1972). Polymorphism, selection and evolution. In *Biochemical Genetics of Man* (D. J. H. Brock and O. Mayo, eds). London: Academic Press.

PEARSON, K. (1904). On the laws of inheritance in man. II. *Biometrika*, **3**, 131.

POLLEY, M. J. and BEARN, A. G. (1974). Cystic fibrosis: current concepts. *J. Med. Gen.*, **11**, 249.

RAO, D. C. and MORTON, N. E. (1974). Path analysis of family resemblance in the presence of gene-environment interaction. *Amer. J. Hum. Genet.*, **26**, 767.

ROBERTS, G. E. (1973). The relevance of biochemistry to the problem of mental retardation. In *Biochemistry and Mental Illness* (L. L. Iversen and S. P. R. Rose, eds). London: The Biochemical Society.

ROBERTSON, A. (1962). Selection for heterozygotes in small populations. *Genetics*, **47**, 1291.

SHEPPARD, W. F. (1897). On the geometrical treatment of the 'Normal Curve' of statistics, with especial reference to correlation and to the theory of error. *Proc. Roy. Soc.*, **62**, 170.

SHIELDS, J., HESTON, L. L. and GOTTESMAN, I. I. (1975). Schizophrenia and the schizoid: the problem for genetic analysis. In *Genetics and Psychopathology* (R. Fieve, H. Brill and D. Rosenthal, eds). Baltimore: Johns Hopkins Press.

SIMPSON, N. E. (1969). Heritabilities of liability to diabetes when sex and age at onset are considered. *Ann. Hum. Genet.*, **32**, 283.

SLATER, E. and COWIE, V. (1971). *The Genetics of Mental Disorders*. London: Oxford University Press.

THODAY, J. M. (1967). New insights into continuous variation. In *Proceedings of the Third International Congress of Human Genetics* (J. F. Crow and J. V. Neel, eds). Baltimore: Johns Hopkins Press.