Ann. Hum. Genet., Lond. (1970), **34**, 85
Printed in Great Britain

Heritability of liability and concordance in monozygous twins

By CHARLES SMITH

Department of Human Genetics, Edinburgh University, Western General Hospital, Edinburgh

Falconer (1965) presented a method of measuring the correlation between relatives for a disease from data on the incidences of the disease in the general population and among relatives of affected individuals. The model assumes an underlying continuous liability to a disease, the liability being made up of many genetic and environmental factors and thus normally distributed. The disease becomes manifest if an individual's liability exceeds a critical threshold level. The derived correlation between relatives may be used, in the absence or elimination of common familial environmental factors, to provide an estimate of the heritability of liability to the disease.

Two sources of bias remain in Falconer's (1965) method of estimation. These arise because affected individuals (those exceeding the threshold) form a truncated group with a skewed distribution. Thus the variance of liability among their relatives may be reduced and the distribution of relatives may be skewed. Falconer (1965) suggested that these biases would be small, but pointed out (Falconer, 1967) that the skewness could bias estimates of heritability of liability in monozygotic (MZ) twins, giving estimates which are too high.

Edwards (1969) overcame the difficulties of skewness and reduced variance by using tetrachoric functions for bivariate normal distributions. He presented a graph which relates population incidence and incidence in relatives of affected individuals for *any* correlation between relatives in liability. His results show that Falconer's original method gives estimates of heritability which are about one-tenth too low. In this paper Edwards's results are confirmed and extended, using a different approach to the problem. Then the genetic interpretation of concordance rates in MZ twins is examined in the light of the theoretical results obtained.

METHODS

The problems of skewness and of reduced variance, inherent in Falconer's method, arise because a truncated (affected) part of the population is chosen. These disappear when the population is considered as a whole.

The model used assumes a normally distributed liability and absence of familial environmental factors, so that the correlation (r) in liability between relatives equals Rh^2, the genetic relationship multiplied by the heritability of liability. The method of solution was to generate an ordered series of genetic classes from a normal distribution. Then from each class, and for different levels of incidence and heritability, the proportion of relatives exceeding the threshold value was estimated. Pooling the proportions for all genetic classes, a set of graphs or tables was derived giving the correlation (or heritability) corresponding to any combination of incidences in the population and in relatives of affected individuals.

The details of this method are shown in Table 1 (p. 88). The incidence (q_P) in the population specifies the threshold point (T). Given the heritability (h^2), the genetic variance and distribution

of liability in the population are known. The frequency (f_i) of the ith genetic class, its mean deviation ($x_i h$) and the mean deviation ($Rx_i h$) of its relatives from the original mean and so from the threshold point (T) can be found. Then taking account of the respective residual variances, the proportions of each genetic class (P_i), and of relatives (P'_i), exceeding the threshold can be estimated. Combining the proportions for all genetic classes, the incidence (q_R) in relatives of affected individuals is obtained.

The various calculations were done by computer using three subroutines to evaluate functions of the normal curve: (1) to get the mean and frequency in each of 40 equally spaced genetic classes (from -4 to $+4$ standard deviation units), (2) to calculate the deviate (x) given the incidence (q), using a complex algorithm from Hastings (1955), and (3) to find the value of q given x, by accumulating the frequency of classes exceeding x. A wide range of values for incidence, heritability and genetic relationship was used.

RESULTS

The results confirm those of Edwards (1969). They are plotted in Fig. 1, to correspond in form with Falconer's original graphs and to allow more accurate readings than are possible from Edwards's paper. The equivalence with Edwards's graph shows that the results depend on the correlation (r) between relatives in liability. This was confirmed in that the same results were always obtained for a given value of r, equal to Rh^2, for different combinations of R and h^2. In the absence of environmental similarities among relatives, the graphs in Fig. 1 are thus general for all coefficients of relationship (R) by choosing the line for r equal to Rh^2. However, the composition of the genetic variance measured by the heritability will differ for different relatives; including all the additive variance but differing proportions of the dominance and epistatic variance for different genetic relationships (e.g. Falconer, 1960).

Incidence in relatives may, by sampling, be less than the population incidence. This would give rise to negative estimates of heritability of liability as shown in Fig. 1. These should be included when pooling estimates, since taking only positive estimates will bias the overall result.

The variance of estimates of the correlation (or heritability) will depend largely on the variance of the estimate of incidence in relatives. This variance, $V(\log_{10} q_R)$, is $0{\cdot}189/A$ on the $\log_{10}$ scale of Fig. 1, where A is the number of affected relatives. Thus the variance of a correlation estimate can also be read simply from Fig. 1. For example, with five affected relatives, the standard error of $\log_{10} q_R$ is $\sqrt{(0{\cdot}189/5)} = 0{\cdot}194$. This is equivalent to 0·95 cm. on the ordinate of Fig. 1, since 4·9 cm. cover one log unit. At $q_P = 0{\cdot}1\,\%$, $1{\cdot}0\,\%$ and $10\,\%$ respectively the standard error of a heritability estimate of 0·5 from first-degree relatives would be, from Fig. 1, approximately 0·10, 0·16 and 0·32, corresponding with the values obtained by Falconer's (1965) formulae.

The simple geometrical pattern of the lines in Fig. 1 provides a good algebraic approximation (to within 5 %) of the values of the correlation (r), namely:

$$\tan\left(\frac{\pi}{4}(1-r)(1+r^5)\right) = \log q_R/\log q_P$$

and solving for r. The fifth-degree term can be omitted unless r exceeds 0·5; that is, when the situation refers to MZ twins for diseases with high heritability.

Another useful expression can adjust the tabulated correlations (or heritabilities) if the patients represent one population (A) and the relatives are drawn from another population (B);

for example, with regard to sex or age. From Fig. 1 an estimate (r_t) is obtained, using the population incidence for B and the incidence of relatives for B. The adjusted correlation is then simply

$$r_t \left\{ \frac{(\log q_{PB}) - 1}{(\log q_{PA}) - 1} \right\}$$

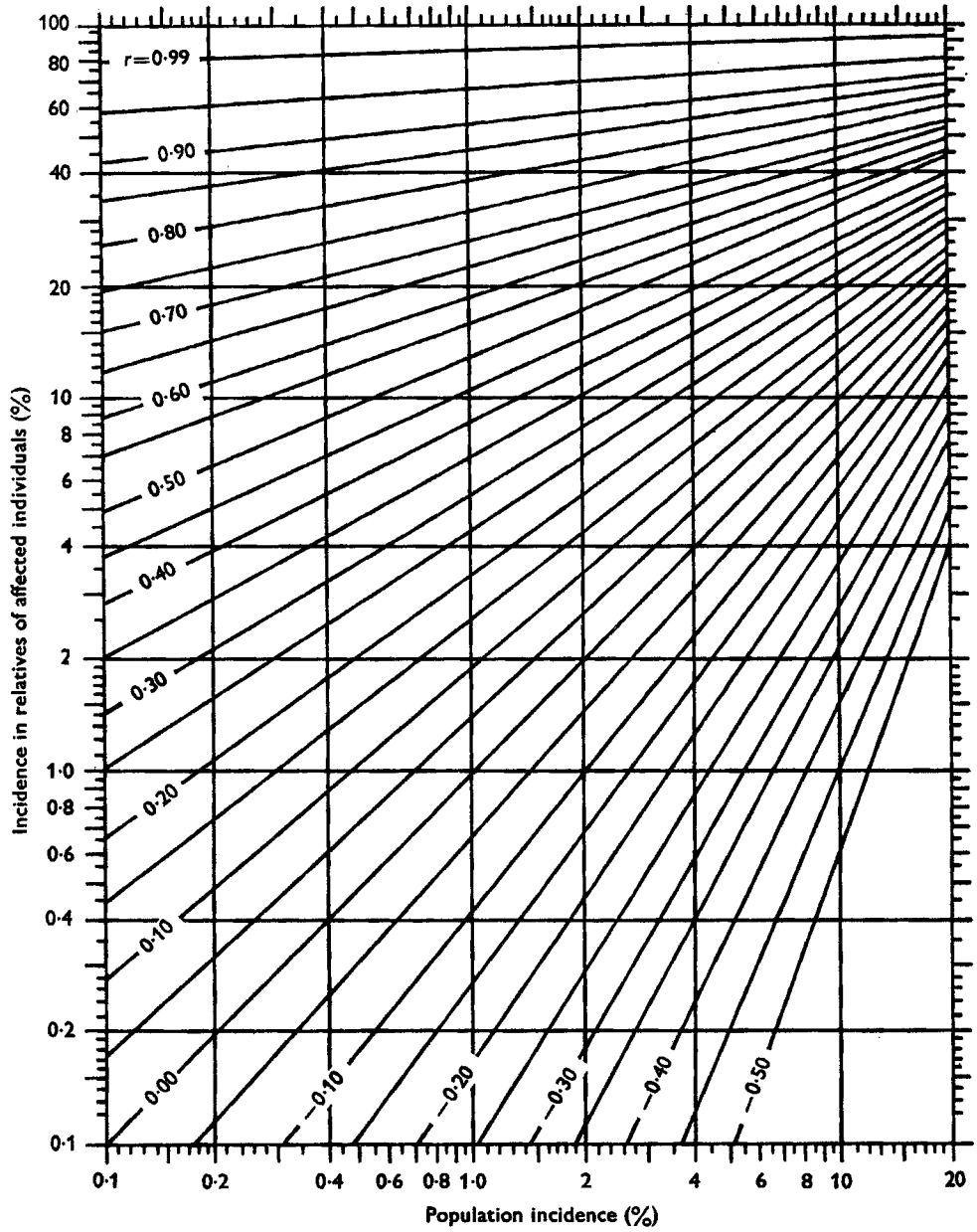

Fig. 1. Graph of the correlation (r) in liability between relatives, given the population incidence and the incidence in relatives of affected individuals. (In the absence of environmental similarities between relatives r equals Rh^2, the coefficient of relationship (R) multiplied by the heritability of liability (h^2)).

To check on the accuracy of this expression, the expected incidence was calculated for relatives in B of patients in A, using the same methods as before and merely changing the threshold level for relatives (Table 1) to that for incidence in B. The expression proved very accurate and could

adjust the correlation (or heritability) for large (up to 100-fold) differences in incidence to within 5% of the true value. For example, if $R = 0{\cdot}5$, $h^2 = 0{\cdot}5$, $q_{PA} = 0{\cdot}1\%$, $q_{PB} = 1{\cdot}0\%$ or 10%, then q_{RB} was found by computer to be 6·2% and 32·5%, respectively, giving the corresponding tabulated heritability values of 0·65 and 1·00. Using the above expression, the adjusted heritabilities became 0·49 and 0·50 respectively.

The results in Fig. 1 also apply to MZ twins, if common environmental effects are absent or can be eliminated. The incidence in co-twins is usually expressed as a concordance rate (the proportion of co-twins affected for each twin independently ascertained) and can be read off Fig. 1 for different levels of heritability and population incidence. To bring out a striking feature of the results, they are shown in another form in Fig. 2. In the absence of environmental similarities, concordance rates in MZ twins will not be expected to be high unless the heritability is very high (or the population incidence is very high). For example, in a disease with incidence of 1% in the general population, the expected concordance rates in MZ twins would only be about 13% if the heritability was 50% or about 37% if the heritability was 80%. Thus a low

Table 1. ***Parameters used in the evaluation of incidence in relatives of affected individuals***

(Population incidence, q_P; threshold, T; heritability, h^2; genetic relationship, R.)

Genetic class (i)	Individuals	Relatives
Genetic mean	$x_i h$	$Rx_i h$
Frequency	f_i	f_i
Residual variance	$1-h^2$	$1-R^2h^2$
Mean deviation from the threshold	$\dfrac{T-x_i h}{\sqrt{(1-h^2)}}$	$\dfrac{T-Rx_i h}{\sqrt{(1-R^2h^2)}}$
Proportion exceeding the threshold	P_i	P_i'
Incidence in relatives of affected individuals	$\sum_i f_i P_i P_i' / \sum_i f_i P_i$	

concordance rate in MZ twins cannot be taken to prove that the heritability is low. High MZ concordance rates may indicate either a very high heritability, or that common environmental factors are important. These alternatives may be resolved, as in conventional twin analysis (e.g. Kempthorne & Osborne, 1961), by comparing estimates of correlations in liability for MZ twins, for dizygotic twins (DZ) and among full-sibs. For example, the expression $2(r_{MZ} - r_{DZ})$ estimates the heritability, while the term $(2r_{DZ} - r_{MZ})$ estimates the importance of common environmental effects in twins.

A good example of the dilemma posed by finding low MZ concordance rates but high heritability estimates from other relatives is given by clubfoot (talipes equinovarus). Ching, Chung & Nemechek (1969) concluded that inheritance was multifactorial with a heritability of liability of 0·68. Wynne-Davies (1970) found an estimate of 0·60 for heritability of clubfoot, using British data. However, they quote, as anomalous with their results, data of Idelberger (1939) on 174 twin pairs (40 MZ and 134 DZ) with at least one member of the pair having clubfoot. The concordance rates were 33% for MZ and 3% of DZ pairs. Taking the population incidence as 0·12% in Caucasians, estimates of heritability of liability from Idelberger's data are:

from MZ	$0{\cdot}83 \pm 0{\cdot}06$,
from DZ	$0{\cdot}78 \pm 0{\cdot}16$,
from $2(r_{MZ} - r_{DZ})$	$0{\cdot}88 \pm 0{\cdot}34$.

Thus the MZ concordance rate is not too low and anomalous with the other data but is as expected (or rather high) for multifactorial inheritance. A similar resolution of apparently conflicting data from twins and from other relatives may be possible for other diseases; for example, for schizophrenia (Gottesman & Shields, 1967).

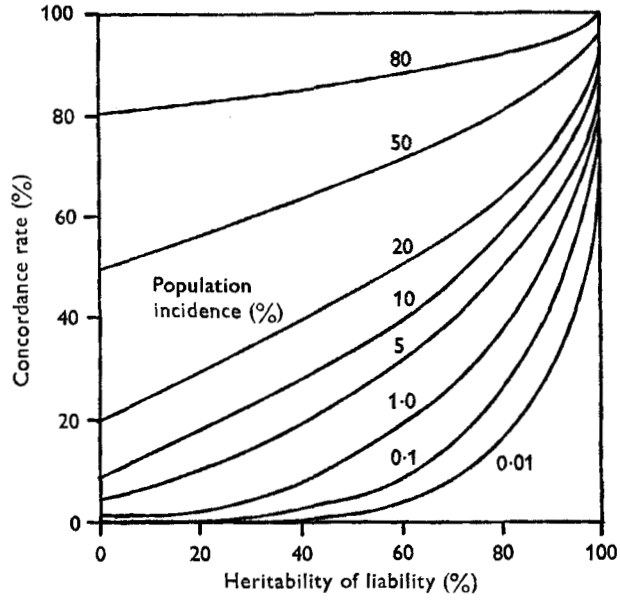

Fig. 2. Expected concordance rate in monozygotic (MZ) twins given the population incidence and the heritability of liability.

DISCUSSION

In comparison with Fig. 1, Falconer's (1965) graph and formulae underestimate the correlation between relatives by about 10% of the true estimate. Thus, in absolute units, the bias is only important when the correlation is high. The anomalous situation of overestimation of heritability in MZ twins, discussed by Falconer (1967), arises only when the correlation between MZ twins is very high (over 0·9). Falconer's (1965) original method may still be useful in many analyses where the present methods and results are difficult to apply—for example, in complex situations involving degree of severity or onset age or any partition of the affected group by some measure of liability.

The abrupt threshold model has been criticized as unrealistic and Edwards (1969) has considered an alternative of attaching a risk function to a normally distributed liability. The empirical risk function used, however, tends to infinity at the limit rather than to unity. The present approach shows how the threshold model implies a normally distributed *genetic* liability, with a cumulative normal risk function, determined by the population incidence (setting the threshold) and by the heritability (determining the residual variance in a genetic class) of disease. The abrupt threshold is thus conceptual rather than real and may be avoided by redefining the variance and risk function.

The improved estimates of heritability of liability should not obscure possible deficiencies of the model in summarizing data on population and familial incidences of a disease. These have been well discussed by Edwards (1969) and others. The method appears to be robust in practice,

and reasonable estimates of heritability have been obtained for several diseases (Carter, 1969). It thus provides a useful tool in understanding and utilizing information on multifactorial disease. However, because its application is simple, there is some danger that the method may be used uncritically and indiscriminately. For example, few authors have tried to estimate or eliminate the effects of environmental similarities among relatives, and these may be important.

A feature of Falconer's (1965) model is that it is descriptive rather than analytical. By combining data on many families genetic heterogeneity may be masked rather than elucidated. However, it may be possible to deduce the underlying genetic basis of a disease—for example, by comparing heritability estimates from different kinds of relatives. Thus, if heritability estimates were larger from sib data than from parent or offspring data, then recessive alleles at one or several loci might be indicated. Some resolution of heterogeneity may also be possible by partitioning a condition into different classes by some criterion. A measure of effectiveness of a partition is given by the genetic correlation between classes, which tends to zero as distinct classes are identified. The heritability of the combined condition will be less (not greater as Edwards (1969) suggests) than the weighted average of the heritabilities in the separate classes. The reverse applies if the partition does not separate different genetic classes within a condition.

Morton *et al.* (1970) have examined the additive continuous models of Falconer (1965) and Edwards (1969), incorporating terms for consanguinity. They have also presented a discontinuous 'load' model which takes account of loci with additive effects and also of rare loci with large non-additive effects. They found that the discontinuous model gave a good fit more often than the other two models when tested on several sets of data, which included incidences in different kinds of relatives and in consanguineous marriages for several familial diseases. However, all models gave similar estimates of recurrence risks.

SUMMARY

Results from heritability of liability model of Falconer (1965) have been revised, eliminating two sources of bias. The results have been graphed for convenient usage, and in the absence of environmental similarities among relatives can apply to all kind of relatives. Confidence levels for the heritability (or correlation) estimates can also be read directly off the graph.

The results also apply to concordance rates in MZ twins. They show that, in the absence of environmental similarities, concordance rates for a disease in MZ twins will only be high if the heritability is *very* high. Thus a low concordance rate in MZ twins cannot be taken to prove that genetic factors are not important in the predisposition to a disease.

I am indebted to Dr O. Mayo and Dr D. S. Falconer for constructive criticism and advice in the development of this paper; and to Dr T. Gedde-Dahl for some unpublished results on concordance in twins. The approach used arose during discussion with Dr W. E. Nance.

REFERENCES

CARTER, C. O. (1969). Genetics of common disorders. *Br. Med. Bull.* **25**, 52–7.

CHING, G. H. S., CHUNG, C. S. & NEMECHEK, R. W. (1969). Genetic and epidemiological studies of clubfoot in Hawaii. Ascertainment and incidence. *Am. J. Hum. Genet.* **21**, 556–80.

EDWARDS, J. H. (1969). Familial predisposition in man. *Br. Med. Bull.* **25**, 58–63.

FALCONER, D. S. (1960). *Introduction to Quantitative Genetics.* Edinburgh: Oliver and Boyd.

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

FALCONER, D. S. (1967). The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann. Hum. Genet., Lond.* **31**, 1–20.

GOTTESMAN, I. I. & SHIELDS, J. (1967). A polygenic theory of schizophrenia. *Proc. Natn. Acad. Sci. U.S.A.* **58**, 199–205.

HASTINGS, C. (1955). *Approximations for Digital Computers*. Princeton, New Jersey: Princeton University Press.

IDELBERGER, K. (1939). Die Ergebnisse der Zwillingsforschung beim angeborenen Klumpfuss. *Verh. dt. orthop. Ges.* **25**, 272–6.

KEMPTHORNE, O. & OSBORNE, R. H. (1961). The interpretation of twin data. *Am. J. Hum. Genet.* **13**, 320–39.

MORTON, M. E., YEE, S., ELSTON, R. C. & LEW, R. (1970). Discontinuity and quasi-continuity. Alternative hypotheses of multifactorial inheritance. (Submitted for publication.)

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