*Journal of Medical Genetics*, 1978, **15**, 106-108

# Use of overlapping normal distributions in genetic counselling

N. R. DENNIS[1] AND C. O. CARTER

*From MRC Clinical Genetics Unit, Institute of Child Health, London*

SUMMARY If a numerical variable can be assumed to be normally, but differently, distributed in each of two or more populations, then for an 'unknown' individual whose numerical value is known the relative probability of his belonging to each of the populations can be simply calculated. We briefly review the method and its application in genetic counselling.

We recently published our experience with the use of creatine kinase (CK) measurements in counselling women who might be carriers of Duchenne muscular dystrophy (Dennis *et al.*, 1976). The simple statistical technique used was not fully reported, but we feel a short note on it here may be useful, since it is widely applicable in clinical medicine, where it is frequently necessary to assign a patient to one of two populations (carrier vs. non-carrier; diseased vs. normal, etc.) by observing or measuring attributes known to have different distributions in the two populations. Such an assignment is never absolutely certain; only more or less probable. It is, however, better to estimate and use such a probability than, as is so often done, to take a single best discriminating value and assign all those with higher values to one group and those with lower values to the other.

Ideally the distribution in each population will be based on a large sample. With smaller numbers it may be helpful if the distribution appears approximately normal to assume normality or to find and use some simple transformation, such as taking logarithms, which will make the distributions appear more normal and then assume normality.

The principle is well known and is covered, for example, by Penrose and Smith in their book *Down's Anomaly* (1966). The mean and standard deviation of the normally distributed variable in each of the two populations must be known or estimated from measurements in individuals known to belong to one or other population. Then from a measurement on an 'unknown' individual the probability of his belonging to each population is proportional to the height of the normal curve for that population at the value measured. This may be derived by converting the individual's value to standard deviation units and looking up the ordinate of the normal curve in tables.

If the populations have different standard deviations a correction is necessary to ensure that the two normal curves represent equal-sized populations; otherwise, the implicit assumption that before the measurement the individual has an equal chance of belonging to either population will not be satisfied. The simplest way of making this correction is to divide the ordinate obtained for each curve by that curve's standard deviation.

The method may be extended to 3 or more populations; the only requirement is that each individual tested must be able to belong to one and only one of them.

Prior information about the probability of an individual's belonging to one or other population can easily be incorporated; the information obtained by the calculations described above then becomes the 'conditional probabilities' of Bayes's theorem (Stevenson and Davison, 1970).

Taking the use of CK measurements in Duchenne muscular dystrophy carrier detection as an example, assume that we have measured CK in an adequate number of control and known carrier women and have found that the logarithms of the CK values follow an approximately normal distribution with mean and standard deviation M, S in the controls and M′, S′ in the carriers. The distributions of CK values (the mean of 3 estimates from each individual woman) in the controls and the obligatory carriers from the original paper of Wilson *et al.* (1965), labelled first series, and since then labelled second series, are shown in the Table. The overlapping parts of the

[1]Present address: Division of Human Genetics, 86 Hodge Avenue, Buffalo, New York 14222, U.S.A.

Received for publication 26 July 1977.

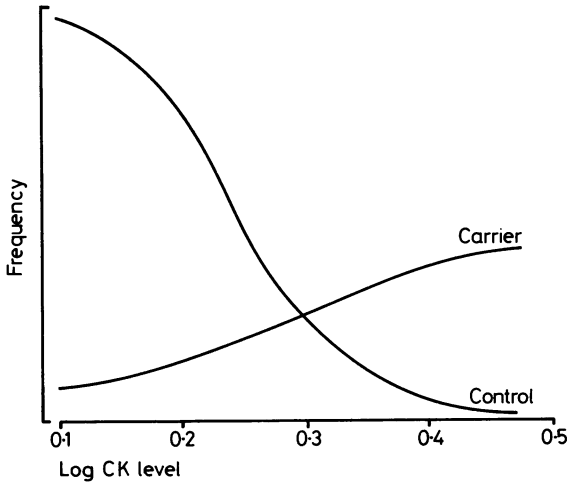


Fig. 1 *Overlapping part of theoretical normal curves derived from log CK distributions in controls and carriers.*

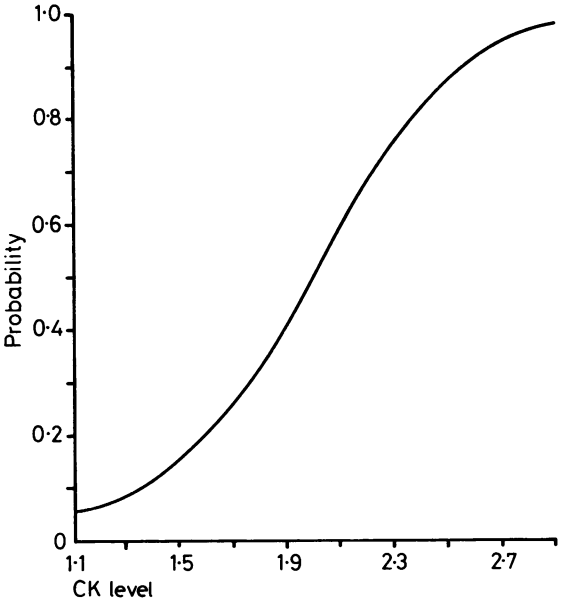


Fig. 2 *Probability p′ (see text), that a woman is a carrier, plotted against CK level.*

theoretical normal curves derived from the combined series of 36 controls and 32 carriers are shown in Fig. 1. The heights of the curves are in inverse proportion to their standard deviations as described above.

A woman whose prior (genetic) probability of being a carrier is 0·25 has a log CK value of X. Then, on the control distribution she lies at a point $(X - M)/S$, with ordinate O, and on the carrier distribution she lies at $(X - M')/S'$, with ordinate O′. The ordinates are divided by the standard deviations of their respective curves, giving relative probabilities of belonging to the control and carrier populations of O/S and O′/S′ respectively. These may be expressed as odds $(O/S):(O'/S')$ or converted to absolute probabilities

$$\frac{O/S}{O/S + O'/S'} \text{ (or p)} \quad \text{and} \quad \frac{O'/S'}{O/S + O'/S'} \text{ (or p′).}$$

Using these as conditional probabilities in Bayes's theorem gives:

$$\text{relative probability of not being a carrier} = 0{\cdot}75\ p$$

$$\text{relative probability of being a carrier} = 0{\cdot}25\ p'$$

$$\text{absolute probability of being a carrier} = \frac{0{\cdot}25\ p'}{0{\cdot}25\ p' + 0{\cdot}75\ p}$$

In practice, if the procedure is being used often it is useful to calculate the probability p′ for various values of X and draw a graph of the type shown in Fig. 2, derived from the data of the Table and the normal curves of Fig. 1. If the measurement being used varies appreciably from time to time in the same person, X should be based on the mean of several measurements, as should the values used in constructing the population distributions.

More than one measurement may be used in this way to characterise two or more populations, provided that the attributes being measured vary independently within the different populations.

Table *CK values (μmol creatine/ml per hr) in 36 control women and 32 known female carriers of Duchenne muscular dystrophy: each value is mean of 3 estimations*

| *Control* | | *Carrier* | |
|---|---|---|---|
| 1·06 | 1·20 | 4·50 | 11·90 |
| 1·07 | 1·83 | 2·53 | 3·60 |
| 0·97 | 1·03 | 3·50 | 5·20 |
| 0·93 | 1·27 | 4·20 | 1·43 |
| 1·53 | 0·92 | 5·23 | 2·10 |
| 1·20 | 0·82 | 3·33 | 2·30 |
| 1·13 | 0·94 | 2·43 | 2·00 |
| 1·27 | 1·85 | 2·80 | 10·20 |
| 0·90 | 1·14 | 1·43 | 4·20 |
| 1·40 | 1·12 | 3·60 | 2·50 |
| 0·80 | 1·09 | 4·63 | 16·90 |
| 1·16 | 1·27 | 10·47 | 3·00 |
| 1·03 | 1·22 | 4·43 | 3·30 |
| 1·77 | 0·93 | 2·13 | 8·90 |
| 1·33 | 2·27 | 3·20 | |
| 1·83 | 1·77 | 5·57 | |
| 2·63 | 1·43 | 4·43 | |
| 1·67 | 1·20 | 2·80 | |

N.R.D. thanks the Dr Henry C. and Bertha H. Buswell Research Foundation, School of Medicine,

State University of New York at Buffalo, for the award of a Buswell Fellowship during the preparation of this paper.

**References**

Dennis, N. R., Evans, K. A., Clayton, B., and Carter, C. O. (1976). Use of creatine kinase for detecting severe X-linked muscular dystrophy carriers. *British Medical Journal*, **2**, 577-579.

Penrose, L. S., and Smith, G. F. (1966). *Down's Anomaly*, pp. 106-107. J. and A. Churchill, London.

Stevenson, A. C., and Davison, B. C. G. (1970). *Genetic Counselling*, p. 72. Heinemann, London.

Wilson, K. M., Evans, K. A., and Carter, C. O. (1965). Creatine kinase levels in women who carry genes for 3 types of muscular dystrophy. *British Medical Journal*, **1**, 750-753.

Requests for reprints to Dr N. R. Dennis, Division of Human Genetics, Children's Hospital, 86 Hodge Avenue, Buffalo, New York 14222, U.S.A.