J. med. Genet. (1964). **1**, 42.

A Quadratic Discriminant of Zygosity from Fingerprints*

PATRICK SLATER, JAMES SHIELDS, and ELIOT SLATER

From the Institute of Psychiatry and the Medical Research Council Psychiatric Genetics Research Unit, Maudsley Hospital, London

Work previously done in this Unit and elsewhere on the use of fingerprints for the diagnosis of zygosity of pairs of twins has been recently reviewed by one of us (Slater, 1963). The author also reported a method for calculating a linear discriminant, which had been empirically derived from data on 180 MZ and 90 DZ pairs of twins whose blood groups were known to be identical in the MZ pairs and different in one or more factors in the DZ pairs. The discriminant, *Z*, was shown to be normally distributed in both MZ and DZ series. The mean value of *Z* in the MZ series was 0·614 with standard deviation 0·219, in the DZ series 1·001 with standard deviation 0·238. Of the 180 MZ pairs 52 had a lower value than any DZ pair, and of the 90 DZ pairs 30 had a higher *Z* value than any MZ pair. The discriminant was validated on an independent series of 48 MZ and 99 DZ pairs, and shown to be a useful tool for discrimination.

However, it seems probable that a linear discriminant is likely to be relatively inefficient; and one that has been obtained by what amounts to guesswork has obvious disadvantages. A more rigorous method of deriving a discriminant is much to be preferred. The fingerprints of our 180 MZ and 90 DZ pairs, supplemented by an additional 4 pairs, have accordingly been re-analysed along the following line.

The problem is to obtain some indices from the 40 measurements provided by the radial and ulnar ridge counts in a pair of twins that will serve to describe the degree of similarity between the twins. The corresponding measurements in the two twins can be tabulated as 20 pairs of counts and represented as a bivariate frequency distribution which can be specified approximately by its means, variances, and correlation, even though it may not conform closely to the normal form (naturally the conformity is closer in some pairs of twins than others). Indices to describe similarity can be derived from these statistics: the means are likely to be closer, the variances more homogeneous, and the correlations higher if the cases paired are MZ twins than if they are DZ. The indices *d*, *l*, and *z* used here are of this nature: *d* is proportional to the difference between the means, *l* is the logarithm of the variance ratio, and *z* is Fisher's transformation of the correlation coefficient.

TABLE I

DZ TWIN PAIR NO. 76

	Digit	Ridge	A	B	Difference
Right hand	1	r	30	23	7
		u	22	16	6
	2	r	18	13	5
		u	19	13	6
	3	r	14	16	−2
		u	0	0	0
	4	r	23	18	5
		u	12	10	2
	5	r	18	18	0
		u	11	0	11
Left hand	1	r	30	22	8
		u	21	13	8
	2	r	19	11	8
		u	20	16	4
	3	r	15	16	−1
		u	0	0	0
	4	r	18	17	1
		u	14	5	9
	5	r	22	19	3
		u	11	0	11
Sum			337	246	91
S. squares			6835	4088	721

Table I shows the set of counts in one pair of twins and the preliminary steps involved in calculating the indices. The twins are taken in the alphabetical order of their first names and the radial and ulnar counts are listed from the first finger (thumb) of the right hand to the fifth finger of the left. In the third column are the differences between the entries in the A and B columns, which we may denote $a-b$. Below are the sums of the terms in the three columns, Σa, Σb, and $\Sigma(a-b)$, and the sum of their squares, Σa^2, Σb^2, and $\Sigma(a-b)^2$. Hence we can obtain,

$$Va = \Sigma a^2 - (\Sigma a)^2/20$$
$$= 6835 - 337^2/20 = 1156{\cdot}55.$$

* Received August 27, 1963.

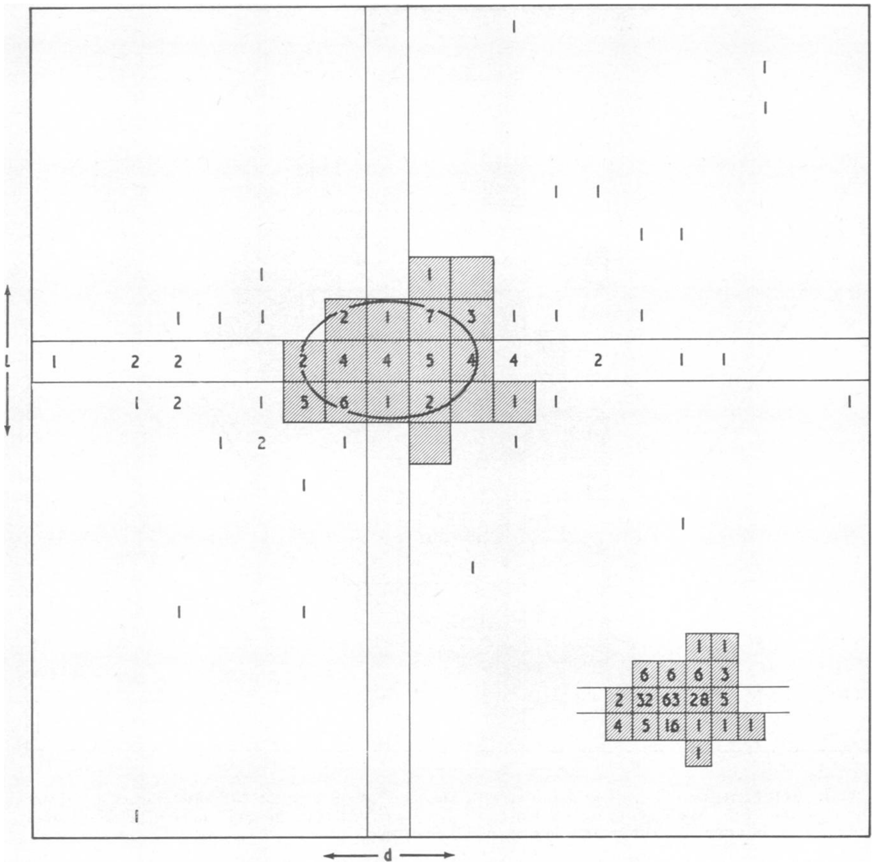

FIG. 1. Scatter of the DZ cases in d and l, with the area occupied by the MZ cases shaded in. Inset shows scatter of the MZ cases in the shaded area. The scale is in units of 25 for d and 0·2 for l from an origin at the mid-point of the central cell.

This is the sum of squares from which the variance of the A counts could be computed. Similarly,

$$Vb = \Sigma b^2 - (\Sigma b)^2/20$$
$$= 4088 - 246^2/20 = 1062{\cdot}20$$

is the sum from which the variance of the B counts would be obtained. Next the covariance is calculated as,

$$Wab = (\Sigma a^2 + \Sigma b^2 - \Sigma(a-b)^2 - 2\,\Sigma a\,\Sigma b/20)/2$$
$$= (6835 + 4088 - 721 - 337 \times 246/10)/2$$
$$= 1911{\cdot}8/2 = 955{\cdot}90,$$

and the correlation is obtained from

$$Wab/(Va\,Vb)^{\frac{1}{2}} = 955{\cdot}90/2455{\cdot}94 = 0{\cdot}3892.$$

The three indices are defined in these terms as

$$d = \Sigma(a - b),$$
$$l = \log_{10}(Va/Vb),$$

and z by Fisher's transformation, which can be read from appropriate statistical tables (Fisher and Yates, 1943, Table VII, or Pearson and Hartley, 1954, Table 14, for example). The values in this case are $d = 91$, $l = 0{\cdot}728$, and $z = 0{\cdot}411$.

Taking the paired cases in the alphabetical order of their first names is a simple randomizing procedure which affects the signs of the values of d and l in a way that seems preferable to alternative procedures such as taking the twin with the larger mean or the larger variance first, or taking all values of d and l as positive by definition. As a result d and l both tend to have zero means and approximately normal distributions in both groups, the difference between the groups being in the range of their variation. A slight tendency appears for d to correlate positively with l, that is to say for the twin with the higher mean ridge count to have the larger variance as well. The correlation is 0·21.

Fig. 1 gives the bivariate frequency distribution separately for each group, monozygotic and dizy-

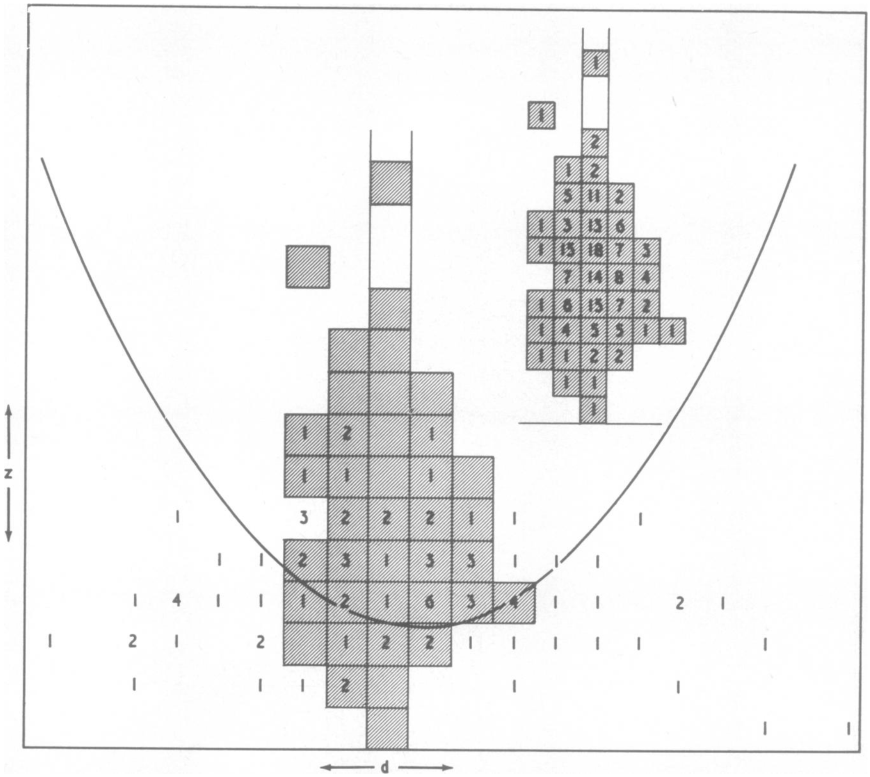

FIG. 2. Scatter of the DZ cases in d and z, with the area occupied by the MZ cases shaded in. Inset shows scatter of the MZ cases in the shaded area. The scale for d is in the same units as before. Values of z increase in units of 0·2 from an origin at the lower margin.

gotic, and shows how the difference manifests itself. Both distributions have their bivariate mean point in the same cell, but the DZ pairs are scattered over the entire region while the MZ pairs are confined to the small shaded area. Their cell frequencies are given in the inset. Approximately 95% are confined within the ellipse. It is clear that the probability of monozygosity increases as the common bivariate mean is approached and decreases towards the periphery of the region: the boundary enclosing the area where the probability exceeds a certain value would presumably be of the elliptical form shown.

The distributions for d and z are given similarly in Fig. 2. Here they are separated by the wide difference in the mean values for z as well as by the difference in the range for d. As before, the shading in the main part of the figure shows the area where the distribution for the MZ pairs would fall if it were superimposed on the other. Again it can be seen that the probability of monozygosity varies over the surface; but here the form of the boundary enclosing the area where it exceeds a certain value would presumably be parabolic. The parabola shown exemplifies the discriminant function proposed.

A figure giving the distributions in l and z would evidently resemble Fig. 2 very closely, and the boundary would presumably appear parabolic in this plane too. Seen in three dimensions, the two dispersions would fit together in a way rather like the yolk and white of a fried egg. To separate the MZ pairs from the DZ something is needed that will act like a scoop: in mathematical language a paraboloid described by a quadratic equation with terms involving d^2, dl, and l^2 as well as d, l, and z.

The individual values of d, l, and z were computed by Miss N. Hemsley and Mrs V. Beevers from the data collected by one of us (J.S.) and referred to the National Physical Laboratory to calculate a discriminant equation in these terms. The work there was supervised by Mrs G. Peters. The same programme is used as for a linear discriminant, the powers of d and l being introduced as additional variables, thus making six in all. The means and variances for the separate groups and

TABLE IIA

GROUP MEANS AND VARIANCES

Variable	MZ pairs (182)		DZ pairs (92)	
	Mean	Variance	Mean	Variance
d	0·2692	447·70	10·5217	8553·06
l	−0·002576	0·011729	−0·010609	0·212864
z	1·1944	0·142592	0·7591	0·090450
d^2	445·31	498795·0	8572·03	165918506·0
dl	0·4881	8·2212	17·0982	3552·60
l^2	0·011671	0·00056120	0·210664	0·389402

TABLE IIB

THE COVARIANCE MATRICES WITHIN GROUPS (MZ ABOVE THE LEADING DIAGONAL; DZ BELOW)

Variable	*d*	*l*	*z*	d^2	*dl*	l^2
d		0·491445	−0·549347	24234·0	0·542210	0·044606
l	17·4001		−0·00126425	0·738840	0·049025	0·00041833
z	−4·3733	−0·011011		−36·2255	0·045887	0·00192907
d^2	461040·0	756·48	−1607·90		213·56	3·2349
dl	479·65	4·4252	−3·6554	230658·0		0·030731
l^2	1·9742	−0·047589	−0·037775	2005·69	27·7180	

their covariance matrices are given in Table II.

The equation for the discriminant, *x*, was computed as

$$x = \begin{bmatrix} 0{\cdot}1577\,d + 5{\cdot}922\,l + 104{\cdot}97\,z \\ -\ 0{\cdot}00459\,d^2 + 0{\cdot}0603\,dl - 26{\cdot}22\,l^2 \end{bmatrix}.$$

Thus in the case of the pair of twins cited in Table I, where $d = 91$, $l = 0{\cdot}728$, and $z = 0{\cdot}411$, the value of *x* is found to be,

$$x = 14{\cdot}35 + 4{\cdot}31 + 43{\cdot}14 - 38{\cdot}01 + 3{\cdot}99 - 13{\cdot}90$$
$$= 13{\cdot}88, \text{ or } 14 \text{ approximately.}$$

It will be found that *d*, *l*, and *z* and their squares and products vary very greatly from case to case, so much so that now one, now another, may be two orders of magnitude greater than any of the others. It is accordingly best to make approximations only when the values in the particular case are known. In general, *x* may be rounded off to a whole number.

The values of *x* in the MZ pairs range from + 7 to + 290, with a median of 124, the distribution being slightly skewed, with its longer tail above the median. The values in the DZ pairs range from + 151 to − 274, with a median of 59; the distribution is very skewed, with its longer tail below the median. Thus the long tails fork away from each other and the short tails are the ones that overlap. Table III shows the frequencies in the region where the overlap occurs.

Within this region both the distributions are approximately normal; and a good fit is obtained by taking the means as 59 for the DZ pairs and 124 for the MZ pairs, and the standard deviation as 40 for both groups ($\chi^2 = 10{\cdot}38$ for the DZ and 5·68 for the MZ, with 13 d.f. each).

So it is reasonable to estimate the probability of monozygosity on the evidence of an observed value of *x* by reference to the properties of the normal frequency distribution. Assuming initial uncertainty, i.e. *a priori* probability of 0·5, the derived probability will depend on the relative value of the ordinate at deviations of $d_m = (124 - x)/40$ and $d_d = (59 - x)/40$ respectively from the mean. Thus the twins considered in Table I, with $x = 14$, have

$d_m = 2{\cdot}750$, ordinate 0·00909
$d_d = 1{\cdot}125$, ordinate 0·21188

Total 0·22097

TABLE III

FREQUENCIES OF MZ AND DZ PAIRS IN SECTIONS OF THE *x*-SCALE WHERE THEIR DISTRIBUTIONS OVERLAP

Scale	Frequencies		Probability of Monozygosity
	MZ	DZ	
160 or over	32	—	0·97
150+	9	2	0·93
140	14	2	0·90
130	18	1	0·85
120	23	1	0·80
110	21	6	0·72
100	14	2	0·63
90	17	6	0·54
80	9	9	0·43
70	10	10	0·34
60	6	7	0·25
50	3	5	0·19
40	3	4	0·13
30	1	4	0·09
20	—	5	0·06
10	1	6	0·04
0	1	2	0·03
Less than 0	—	20	0·01

(c.f. 2, Table I), and the derived probabilities are

$$MZ\ 0{\cdot}00909/0{\cdot}22097 = 0{\cdot}04$$
$$DZ\ 0{\cdot}21188/0{\cdot}22097 = 0{\cdot}96,$$

i.e. the value of x gives a reasonable degree of confidence that the twins are dizygotic.

Approximate values of these probabilities for various values of x in the region of practical interest are listed in the last column of Table III.

Although this method of analysis must come very close to exhausting all the information relevant to the problem of discrimination that can be extracted from the data, the results leave it open to doubt whether it is appreciably superior for practical purposes to the earlier method proposed by Slater (1963). This suggests that the earlier method succeeded in fact in coming close to extracting all the relevant information, although it lacks a convincing rationale. It also suggests that there may be an irreducible area of uncertainty in the application of fingerprint analysis for zygosity diagnosis; that in fact there is a significant proportion of DZ twins that resemble one another as much as the majority of MZ twins. Such a situation might arise if there were only a few independent genetical factors involved, and if a large part of the variance were determined by non-genetical or random factors. This latter suggestion is also supported by the considerable differences in fingerprint formulae commonly found between right and left sides.

The two methods now proposed might be tried out side by side to see whether they duplicate or supplement one another.

Summary

The 20 pairs of ridge counts from the fingerprints of a pair of twins form a bivariate frequency distribution described efficiently by its means, variances, and correlation. From these parameters three indices can be derived to express the degree of similarity between the twins, as far as revealed by the counts.

An examination of the distributions of these indices obtained from two samples, of 182 MZ and 92 DZ twins respectively (classified by blood tests), suggests that a quadratic discriminant function of them is appropriate when they are to be used to determine zygosity. Such a function was accordingly computed with the assistance of the National Physical Laboratory.

It turns out not to give appreciably better results than a formula already reached empirically by Slater (1963). This suggests that there are sources of random variation limiting the reliability of evidence from ridge mounts, and that both the proposed procedures converge on the limit.

References

Fisher, R. A., and Yates, F. (1943). *Statistical Tables for Biological, Agricultural and Medical Research*, 2nd ed. Oliver and Boyd, London.

Pearson, E. S., and Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, Vol. 1. Cambridge University Press, London.

Slater, E. (1963). Diagnosis of zygosity by fingerprints. *Acta psychiat. scand.*, **39**, 78.