THE RESEMBLANCE BETWEEN TWINS, A STATISTICAL EXAMINATION OF LAUTERBACH'S MEASUREMENTS

R. A. FISHER

Rothamsted Experimental Station, Harpenden, Herts, England

Received October 9, 1925

TABLE OF CONTENTS

INTRODUCTORY

In 1919 the author called attention (FISHER 1919) to the remarkable conclusions which appeared to flow from the measurements of twins obtained by THORNDIKE (1905) from the New York schools. At that time this was the only considerable body of measurements available, and although physical measurements were given for only 39 pairs, of which 8 were of unlike sex, the discrepancy between the results and those of the accepted biological theory was so sharp that its statistical significance was not in doubt. The main points of discrepancy were as follows:

(a) The twins of unlike sex showed an average correlation of 0.78, a value much above the value 0.50 to be expected from children related fraternally; since the average of all twins was 0.85, it appeared that twins of unlike sex were not appreciably less alike than twins of like sex. This fact, if it stood alone, would merit little statistical weight, since only eight pairs of unlike sex were available.

(b) The degree of resemblance between each pair in each particular measurement agreed in its curve of distribution remarkably closely with the curve to be expected theoretically from homogeneous material. If some twins differed from others in their closeness of genetic kinship, it was to be expected that this heterogeneity should make the group of values more variable than it was actually found to be.

(c) Twins chosen as especially alike in one trait were found, when other traits were compared, to show no more than the average degree of resemblance. THORNDIKE had not failed to remark this surprising fact, which he described by the term "specialization of resemblance." The correlation between the measure of resemblance of the same pair of twins in different traits, was found in THORNDIKE'S material to be -0.016 with a probable error of only ± 0.028, suggesting that it was in fact zero, and not distinctly positive as was to be expected.

It appeared at the time that there was no alternative but to accept the facts as THORNDIKE had found them, and to attempt to modify the accepted theory. The valuable and extensive data obtained by Professor LAUTERBACH (see foregoing article) seem to render any such modification unnecessary, and to provide for the first time a quantitative demonstration of the different degrees of resemblance between twins of different physiological origin.

It is not indeed possible from an examination of the measurements to divide a group of twins into two distinct classes as "fraternal" and "identical," and so long as this is the case it were perhaps premature to say that all is plain sailing with the theory of twins. The main difficulties, however, raised by THORNDIKE'S data are replaced in LAUTERBACH'S much more extensive series of measurements by positive confirmation of the accepted theory of uniovular and biovular twins.

CORRELATION BETWEEN TWINS OF UNLIKE SEX

LAUTERBACH'S data are sufficiently extensive to determine the correlation between twins of unlike sex with some accuracy. Of the 63 cases, 53 are complete for the four traits, stature, stem length, weight and cephalic index. The great difficulty is that all the pairs are of different ages, and that for the first three traits the growth over the age interval concerned, 7.5 to 19 years, is far from uniform. All three traits show a well marked maximum growth rate, the age of most rapid growth being about 2 years earlier in girls than in boys. The changing growth rate renders the partial correlation eliminating age almost meaningless. A procedure which gets over this difficulty is to fit cubic regression formulae, respectively, to the whole group of boy twins and to the whole group of girls. The deviations from such regression formulae may, as I have shown elsewhere (FISHER 1924), be treated as homogeneous deviates in order to obtain a valid estimate of the correlation. It is still possible that the correlation for twins of mixed sex will be less than that of fraternal twins of the same sex, for factors affecting early or late maturity will inevitably, during the growing period, act less similarly upon twins of unlike sex than on those of like sex; nevertheless, the procedure should be adequate to give a definite answer to the question whether twins of unlike sex are or are not more alike than ordinary brothers and sisters are known to be.

The sums of squares and sums of products of these deviations are shown in table 1.

TABLE 1

Correlation in twins of unlike sex.

	SUM OF SQUARES		SUM OF PRODUCTS	CORRELATION COEFFICIENT	NUMBER OF CASES
	Boys	Girls			
Stature (mm)......	329,947	358,500	156,441	.4549	63
Stem length (mm)..	50,204	84,206	29.979	.4611	53
Weight (lb)........	14,527	13,813	5,386	.3802	62
Cephalic index.....	......	......	......	.5370	63

The cephalic index shows no age or sex differentiation; the mean value was 806.47 and the variance 1339, the correlation for the boy-girl twins being that given in table 1.

It is apparent that the values agree well with the value to be expected if twins of unlike sex are related in the same manner as ordinary brothers and sisters; the average correlation, 0.4583, with a standard error about ± 0.053, agrees sufficiently well with the usual values, about 0.50, even if we make no allowance for the possibility that the correlation in weight is not really so high as are the correlations in the skeletal measurements.

The contrast with THORNDIKE'S data is not due to difference in treatment. If we take out individual measures of resemblance as was done for THORNDIKE'S data, we obtain the median values for stature 0.585, for stem length 0.405, for weight 0.474, and for cephalic index 0.454, the average being 0.480 and the median of all 241 values being 0.500. These latter estimates are not so accurate as those obtained above, since they depend only on the ratios of the deviations, and ignore their actual values; they simply demonstrate that the two methods yield concordant results. We thus have in these data the first experimental verification of the belief that twins of unlike sex show the same degree of resemblance as do ordinary brothers and sisters.

AVERAGE CORRELATION BETWEEN TWINS OF LIKE SEX

For twins of like sex both measurements will be referred to the same standard, representing the mean measurement for that age in the sex concerned; we may, therefore, use HARRIS'S (1909) abbreviated method of calculation of intraclass correlations. We take the difference in the measurements for each pair of twins, and find the mean square of these differences; dividing this by the mean variance at fixed age, we have $2(1-r)$, where r is a trustworthy estimate of the correlation. In these intraclass correlations the effect of age is only involved in the estimates of the mean variance at fixed age.

TABLE 2

The mean square differences between twins of like sex.

	STATURE IN MILLIMETERS	STEM LENGTH IN MILLIMETERS	WEIGHT IN FOUR-OUNCE UNITS	CEPHALIC INDEX
Boys...........	3316	1492	2261	904
Girls...........	2355	863	1358	542
Together.......	2826	1180	1800	720

TABLE 3

The mean variances at fixed age in twins of like sex.

	STATURE IN MILLIMETERS	STEM LENGTH IN MILLIMETERS	WEIGHT IN FOUR-OUNCE UNITS	CEPHALIC INDEX
Boys...........	6068	1314	3134	1313
Girls...........	4631	1463	3644	1365
Together.......	5339	1388	3393	1339

TABLE 4

Correlations in twins of like sex, derived from the data of tables 2 and 3.

	STATURE	STEM LENGTH	WEIGHT	CEPHALIC INDEX
Boys...........	.7268	.4323	.6393	.6558
Girls...........	.7457	.6891	.8137	.8015
Together.......	.7353	.5749	.7347	.7310

For stature, weight and cephalic index the correlations, 0.73 to 0.74, are very substantially higher than for twins of unlike sex; for stem length the values, especially that for boys, show no such considerable difference. In this connection it should be mentioned that if the like-sex twins are of two kinds, showing, respectively, "fraternal" and "identical" degrees of resemblance, the mean square differences will be dominated by the former group and the random-sampling errors will be very considerably greater than with homogeneous material.

It is thus possible to ascribe the lower values for stem length to a few exceptionally large unlike deviations among fraternal twins of like sex.

On the other hand, it is definitely impossible to regard the like-sex twins as homogeneous in respect of degree of resemblance with the twins of unlike sex; for in this case the correlations for like-sex twins would not differ from those of twins of unlike sex by more than the random-sampling errors to be expected from homogeneous material. These errors have been studied in detail; the standard error of the correlation coefficient as ordinarily quoted does not provide a reliable test, but, as I have explained elsewhere (FISHER 1925), an accurate test is possible by means of the related quantity, z, which may be regarded as a transformal correlation.

For stature the values of z are compared in table 5.

TABLE 5

Values of z for statures of like- and unlike-sex twins.

	r	z	RANDOM-SAMPLING VARIANCE
Like-sex	.1353	.9402	.006969
Unlike-sex	.4549	.4909	.016667
Difference		+.4493	.023636

The difference in z is 0.4493 with a standard error ± 0.1537. Treating the other values similarly we have the results shown in table 6.

TABLE 6

Differences of like- and unlike-sex twins with respect to the quantity, z.

	STATURE	STEM LENGTH	WEIGHT	CEPHALIC INDEX
Difference in z	+.45	+.16	+.54	+.33
Standard error	±.15	±.17	±.15	±.15

All save stem length show significant differences, and in stem length the difference, though not significant is in the same direction as the others. It is thus obvious that the like-sex twins do not form material homogeneous with those of unlike-sex, but that some or all of them are much more highly correlated. The correlations from LAUTERBACH'S data are considerably lower than those from THORNDIKE'S data ($r=0.80$), even when we include in the latter one case in five of unlike sex. This suggests that THORNDIKE encountered a considerably higher proportion of identical twins, in addition to a group of twins of unlike sex with unusually close resemblance.

HETEROGENEITY OF TWINS OF LIKE SEX

In view of THORNDIKE'S data which showed no heterogeneity among twins of like sex, it is important to determine if these data show heterogeneity. The most direct test depends on the differences in the measurements of like-sex twins. If d is the difference of any one pair, found by subtracting the less measurement from the greater, and $\bar{d}$ stand for the mean difference, $\overline{d^2}$ for the mean of the squared difference, then for a large sample of normally distributed values we should have

$$\overline{d^2} = \frac{\pi}{2}\bar{d}^2$$

whereas, for a mixture of two such populations, with different mean differences, $\overline{d^2} - \frac{\pi}{2}\bar{d}^2$ should be positive. To utilize this fact it is necessary to know the standard error of $\overline{d^2} - \frac{\pi}{2}\bar{d}^2$ and this is found to be

$$\frac{\overline{d^2}}{\sqrt{n}}(2\pi - 6) = \frac{\bar{d}^2}{\sqrt{n}} \times .5321$$

Applying this test to LAUTERBACH'S data, I find the values given in table 7.

TABLE 7

Tests for heterogeneity in like-sex twins.

	STATURE IN MILLIMETERS		STEM LENGTH IN MILLIMETERS		WEIGHT IN FOUR-OUNCE UNITS		CEPHALIC INDEX PER MILLE	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
n	71	74	66	65	70	73	68	70
$\bar{d}$	40.479	34.622	26.273	17.923	33.286	26.260	22.544	17.410
$\bar{d}^2$	3316	2355	1492	863	2261	1358	904	542
$\overline{d^2} - \frac{\pi}{2}\bar{d}^2$	+742	+472	+408	+359	+521	+274	+106	+66
S.E.	±209	±146	±98	±57	±144	±85	±58	±34

In all cases there appears to be significant evidence of heterogeneity; in the case of cephalic index the separate values for boys and girls are scarcely significant, but taking boys and girls together they provide significant evidence.

Is this apparent heterogeneity due to heterogeneity of origin? At first sight other causes cannot be excluded. The data are admittedly heterogeneous in respect of age, and the fact that we are using deviations from fitted regression curves on age will doubtless introduce some further heterogeneity. The effect ascribable to these two causes may perhaps best be estimated by treating the twins of unlike sex in the same manner. For these the results are given in table 8.

TABLE 8

Tests for heterogeneity in unlike-sex twins.

	STATURE IN MILLIMETERS	STEM LENGTH IN MILLIMETERS	WEIGHT IN FOUR-OUNCE UNITS	CEPHALIC INDEX PER MILLE
n	63	53	62	63
$\bar{d}$	60.048	30.434	48.852	30.937
$\overline{d^2}$	5961	1406	4462	1348
$\overline{d^2}-\frac{\pi}{2}\bar{d}^2$	297	−49	713	−155
S.E.	±400	±103	±301	±90

Two of the deviations are positive and two negative; only that for weight is significant, and this is just the trait in which the effect of age heterogeneity should be most marked. The evidence speaks decisively in favor of the view that twins of like sex are heterogeneous in their mode of origin, while those of unlike sex are apparently homogeneous.

On the theory of fraternal and identical twins about 40 percent of those of like sex should be identical. We cannot assume this ratio *a priori* for the particular sample measured by LAUTERBACH. The correlations indicate, for example, that a larger proportion of the girl twins are identical than of the boys.

The proportion identical in the whole group of boys and girls, must be nearly the same for all measurements, and would be absolutely the same if all cases had been completely measured; if we assume that the standard difference is the same for fraternal twins of like sex as it is for twins of unlike sex we may use the figures for cephalic index to estimate the proportion of identical twins present. If a is the standard difference for identical twins, and p the proportion identical, we shall have

$$pa+(1-p)36.715 = 24.997$$
$$pa^2+(1-p)1348 = 720.38$$

whence,

$$a = 16.856$$
$$p = .59007$$

If we adopt the value 59 percent identical, we may infer the correlation in the identical group from the actual correlations of the like-sex pairs, taking the average value for the unlike-sex pairs, 0.4583, to represent the fraternal correlation; in this way we obtain

	Stature	*Stem length*	*Weight*	*Cephalic index*
Correlation for identical twins	.9278	.6898	.9268	.9205

Apart from stem length the agreement is excellent.

The values obtained for the estimated correlation between identical twins are of great interest, since, if they could be determined with certainty, they would afford a direct measure of the importance of genetic factors in determining these traits. From an examination of PEARSON'S data for the correlation of related adults the author concluded (FISHER 1918) that, if heredity is due to a number of cumulative Mendelian factors, more than 90 percent of the variance must be due to genetic causes. PEARSON'S data referred to stature, span and cubit, and LAUTERBACH'S data not only confirm the fact for stature, but make it probable that it should be extended to cephalic index, and, for children at least, to weight. It should be remembered that the differences observed in these identical twins are absolutely small. The standard difference in stature is only 26 mm, and from my own experience of measuring children a standard error of measurement of 3 to 5 mm would not seem excessive. The effect of random errors of measurement will be to lower the correlations and will be especially important for stem length, when the measurement errors are at least as great as for stature, while the absolute differences are much less. For cephalic index the standard difference is only 17 parts per mille, and an error of 1 mm in head breadth will produce an error of nearly 6 parts per mille. The average of the four correlations as estimated for identical twins, 0.894, will therefore have been subjected to a dilution of uncertain amount due to errors of measurement.

SPECIALIZATION OF RESEMBLANCE

THORNDIKE'S data clearly indicated that twins more alike in one character were not, on the whole, more alike in other characters. This conclusion is so contrary to the accepted hypothesis that, if it were substantiated, that hypothesis must be abandoned.

In the present data three of the characters, height, stem length and weight, are somewhat closely associated together, and the best test of specialization will be to compare resemblance in stature and cephalic index.

The like-sex pairs were divided according as the difference in stature was 0 to 52 mm, or 53 mm upwards, and also according as the difference in cephalic index was 0 to 25, or 26 upwards. The results are given in table 9.

TABLE 9

Boys

Difference in cephalic index	Difference in stature: Low	High	Total
Low	33	10	43
High	20	5	25
Total	53	15	68

Girls

Difference in cephalic index	Difference in stature: Low	High	Total
Low	40	11	51
High	13	6	19
Total	53	17	70

Neither table, nor the two thrown together, gives any indication that those more alike in stature are more alike in cephalic index, or *vice versa*. To test the matter more exactly, the mean square difference in stature and cephalic index was found for each class (table 10).

TABLE 10

Mean square difference in stature.

Boys

Difference in cephalic index	Difference in stature: Low	High	Total
Low	598	13165	3521
High	869	11911	3077
Total	700	12747	3358

Girls

Difference in cephalic index	Difference in stature: Low	High	Total
Low	444	9634	2426
High	899	6218	2578
Total	555	8428	2467

With boys, those differing greatly in cephalic index have on the average a somewhat less difference in stature, than those with similar cephalic indices; with girls, the reverse is true, but to a less extent. If, however, we confine attention to those with low differences in stature, so excluding the high values which tend to dominate the averages, both sexes show a distinctly lower difference in stature, for the children with similar cephalic indices. Putting the two sexes together we have

	Cases	*Mean square*
Like in cephalic index	73	516
Unlike in cephalic index	33	881

This difference is just over the verge of significance; by my z test (FISHER 1925, chapter VII) I find $z=0.267$, while the value 0.234 is exceeded by chance in only 5 percent of cases.

Repeating the test for differences in cephalic index, we have the grouping shown in table 11. Again, in the totals, the boys alike in stature show actually greater divergence in the cephalic index, while the girls show the

TABLE 11

Mean square difference in cephalic index.

Boys

Difference in cephalic index	Difference in stature: Low	High	Total
Low.....	156	203	167
High.....	2382	1296	2173
Total....	1000	568	904

Girls

Difference in cephalic index	Difference in stature: Low	High	Total
Low.....	122	227	.145
High.....	1494	1853	1607
Total....	458	801	542

reverse to a less extent. Confining the comparison to those alike in cephalic index, the two sexes agree, and give

	Cases	*Mean square*
Like in stature....................	73	137
Unlike in stature..................	21	216

By the z test, $z=0.228$, and the 5 percent point is at 0.266.

Both results, therefore, are on the verge of significance, and together seem to establish that, if we exclude the cases of wide discrepancy, twins much alike in head form tend to be more alike in stature, and *vice versa*.

CONCLUSIONS

LAUTERBACH'S physical measurements of about 200 pairs of twins appear to provide unequivocal evidence for the following conclusions:

(1) Twins of unlike sex resemble each other to approximately the same extent as do ordinary brothers and sisters.

(2) Twins of like sex show on the average a considerably closer resemblance.

(3) Twins of like sex are heterogeneous, and are therefore divisible in respect of resemblance into at least two classes.

(4) The data may be interpreted as due to a mixture of identical and fraternal twins, of which about 59 percent appear to be identical. The correlations between identical twins must, on this supposition, be about 0.9 or over.

(5) If we set aside twins with large differences in stature as certainly fraternal, the remainder show that those with large differences in cephalic index have on the average larger differences in stature; *mutatis mutandis*, the same is true of cephalic index. The data thus supply, for the first time, evidence of association of resemblance in different traits.

LITERATURE CITED

FISHER, R. A., 1918 The Correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh **52**: 399-433.

1919 The genesis of twins. Genetics **4**: 489-499.

1924 The influence of rainfall on the yield of wheat at Rothamsted. Phil. Trans. Roy. Soc. London **213**: 89-142.

1925 Statistical methods for research workers. 239 pp. Edinburgh: Oliver and Boyd.

HARRIS, J. A., 1909 A short method of calculating the coefficient of correlation in the case of integral variates. Biometrika **7**: 214–218.

THORNDIKE, E. L., 1905 Measurements of twins. 64 pp. New York: Science Press.