Reproduced with permission of Blackwell Science

253

THE VARIATION IN STRENGTH OF THE HUMAN BLOOD GROUP *P*

SIR RONALD FISHER
Department of Genetics, Cambridge

Received 4.xii.52

1. INTRODUCTION

Soon after their discovery (1) of the blood group known by the genetical symbols *P*, *p*, Landsteiner and Levine called attention to the great variability of the *P* reaction with anti-*P* sera. This variability has proved very troublesome, and for some years the validity of estimates of gene frequencies was in some doubt, owing to the possibility that an unknown proportion of weak positive reactions were passing as negative. Moreover, there is still no anti-*p* reagent to supplement the typing antibody first discovered.

Owing to the improved technique developed by Jonsson and Henningsen (2), and the use of potent testing fluids, the difficulty of distinguishing weakly reacting blood containing *P* from the recessive form *pp* is now believed to be overcome. The causes of variation in strength are, however, still obscure. Henningsen has indeed suggested (3) that his data may be interpreted as showing a series of four alleles :—

p	47 per cent.
P weak	10 ,,
P medium	40 ,,
P strong	3 ,,
	100 ,,

After examining the data on which this suggestion was based, I suggested in correspondence that a more detailed examination of this fine body of material might be profitable. Dr Henningsen, therefore, very cordially sent me further details, and some additional observations, together with information as to factors such as age of the donor, freshness of the sample, etc., which have guided me in the choice of the further calculations I have made. Especially because my conclusions diverge somewhat from those to which Henningsen was first led, I should like to make it clear that my enquiry rests exclusively on his work, and has been made possible only by his courtesy.

2. THE FAMILY DATA

In the case of 221 families both parents were classified in four phenotypic classes as Strong, Medium, Weak and Negative. As the frequency distribution for children differed considerably from that

 Heredity, 7: 81-89, (1953).

for the parents, I have felt on safe ground in recognising for these only the distinction of Positive and Negative. This distinction is sufficient to allow of an estimation of limited but useful precision of the proportion of each phenotypic class of parent who are heterozygous, meaning by this that they contain the negative gene p, whatever allelomorph it may be combined with.

For this purpose we need to know for each of the nine phenotypic classes of mating, which might show segregation among the children : (i) How many matings have yielded at least one negative child, (ii) in the case of matings with no negative child, how many children have been tested. The number of children in families of the first class is a matter of indifference, since any one negative child proves both parents to be heterozygous, if not known to be themselves negative, and further information can tell us no more than this. In the cases of families with no negative children on the contrary, the larger the family the higher is the probability that at least one parent is homozygous.

With these considerations in view, the whole of the data may be summarised compactly as in table 1.

TABLE 1

Type of mating	Numbers of matings with negative children	Numbers of tested children in matings with no negative children	Numbers of matings with no negative children	Total numbers of matings
Strong—Strong	3	$1^2\,2^5\,3\,4^2\,6$	11	14
Strong—Medium	2	$1^8\,2^{10}\,3^8\,4^5\,5\,6^2\,13$	35	37
Strong—Weak	0	$1^2\,2^3\,3^2$	7	7
Strong—Negative	5	$1^5\,2^3\,3^3\,4^2\,5^3$	16	21
Medium—Medium	16	$1^9\,2^9\,3^{10}\,4^4\,5^3\,6\,7$	37	53
Medium—Weak	6	$1^2\,2^6\,3^3\,4^3\,6^2$	16	22
Medium—Negative	23	$1^8\,2^8\,3\,4\,8,\,10$	20	43
Weak—Weak	1	2	1	2
Weak—Negative	8	1	1	9
Negative—Negative	13	...	0	13
Total	77	...	144	221

Of the total 442 parents tested 99 were negative. The proportion of recessive genes is estimated to be

$$\sqrt{99/442} = 47{\cdot}3267 \text{ per cent.}$$

Consequently, the frequency of positive genes of all possible sorts is 52·6733 per cent., and the proportion of persons with two such genes, who may be called homozygous positives, is

$$(52{\cdot}6733 \text{ per cent.})^2 = 27{\cdot}745 \text{ per cent.}$$

We may, therefore, using the proportion of negative parents only, estimate the number of homozygous positives to be

$$442 \times 27{\cdot}745 \text{ per cent.} = 122{\cdot}6.$$

These must be distributed in some way among the 93 strong, 208 medium and 42 weak recorded. If estimates of the numbers of homozygotes in each of these classes give a concordant estimated total, without using the number of parents negative, this will be a genuine confirmation of the consistency of the data, and of the appropriateness of the method of interpretation.

3. THEORY OF THE CALCULATIONS

If the probability of each child being positive is 3/4, the probability that none will be negative out of s is $(3/4)^s$, and the probability of one at least being negative is

$$u_s = 1-(3/4)^s.$$

Similarly, if, as in a backcross, the probability for each child is 1/2, the probability of at least one negative child is

$$v_s = 1-(1/2)^s.$$

These auxiliary quantities, so far as they are needed for the actual family sizes recorded are shown in table 2.

TABLE 2

Values of u_s *and* v_s *for various sizes of family*

Size of family	u_s $1-(3/4)^s$	Size of family	v_s $1-(1/2)^s$
1	0·25	1	0·5
2	0·4375	2	0·75
3	0·578125	3	0·875
4	0·683594	4	0·9375
5	0·762695	5	0·96875
6	0·822021	8	0·996094
7	0·869680	10	0·999024
13	0·976243		

If p_1 is the proportion of heterozygotes among individuals classified as Strong, p_2 the proportion among the Medium, and p_3 among the Weak, the proportion of the various mating types capable of giving Negative children can be easily calculated. *E.g.* in the mating type Strong mated to Strong, the proportion of matings in which both parents are heterozygous will be

$$p_1^2,$$

and the probability of a family of s having at least one Negative child will be

$$p_1^2 u_s,$$

while the probability of such a family having no Negative child will be

$$1-p_1^2 u_s.$$

Efficient scores for families of s children of these two kinds are found by taking the logarithms, and differentiating with respect to the unknown parameter p_1. Thus we have

Score for family with at least one Negative child $\quad 2/p_1$

Score for family with no Negative child $\quad \dfrac{-2p_1u_s}{(1-p_1^2u_s)}$

In connection with these scores we shall also be concerned to evaluate the amount of information concerning p_1 supplied by the family in question, and this may be evaluated from the differential coefficient of the score with respect to p_1, taken with reversed sign.

So, for a family of s children from a mating with both parents Strong, we have

From a family with at least one Negative child, information

$$2/p_1^2 = S/p_1$$

From a family with no Negative child $\quad -S/p_1+S^2$.

For a mating of Strong and Medium parents, we have correspondingly the probability

$$p_1p_2u_s$$

of at least one Negative child, and

$$1-p_1p_2u_s$$

of no Negative child. Such families are to be scored both for p_1 and for p_2, the scores being

	For p_1	For p_2
Negative child	$1/p_1$	$1/p_2$
No Negative child	$-p_2u_s/(1-p_1p_2u_s)$	$-p_1u_s/(1-p_1p_2u_s)$

From these again the amounts of information are S_1/p_1 and S_2/p_2 for families with a Negative child, while in the alternative case they are

$$S_1^2, \qquad S_1S_2, \qquad S_2^2$$

the middle term being the cross information for the simultaneous estimation of p_1 and p_2.

A third case is offered by the mating Strong with Negative. Here the probabilities are

	p_1v_s	and	$1-p_1v_s$
with scores	$1/p_1$	and	$v_s/(1-p_1v_s)$

and amounts of information

S/p_1	and	S^2

respectively.

Since, on the whole, parents classified as Weak produce Negative children more rather than less frequently than would be expected if

all were heterozygous, we shall simplify the calculations by taking $p_3 = 1$. In this case the matings Strong by Weak and Medium by Weak will be scored just as are Strong by Negative and Medium by Negative, save that u_s will now be used in place of v_s.

4. THE ESTIMATION OF THE PROPORTIONS HOMOZYGOUS

Table 3 gives in detail the scores and amounts of information credited to each class of family according to the type of mating involved. The trial values employed, 1/3 for p_1, and 0·7 for p_2, were known to be very near the values of maximum likelihood. The total scores for p_1 and p_2 are found to be

$$0{\cdot}4383 \quad \text{and} \quad 0{\cdot}5286.$$

The totals for all families also gives the information matrix

$$\begin{Bmatrix} 158{\cdot}4093 & 11{\cdot}9189 \\ 11{\cdot}9189 & 306{\cdot}3893 \end{Bmatrix},$$

so that adjustments δp_1 and δp_2 to the trial value may be calculated from the equations,

$$\begin{aligned} 158{\cdot}4093\,\delta p_1 + 11{\cdot}9189\,\delta p_2 &= 0{\cdot}4383 \\ 11{\cdot}9189\,\delta p_1 + 306{\cdot}3893\,\delta p_2 &= 0{\cdot}5286. \end{aligned}$$

The covariance matrix is found by inverting the information matrix, and multiplied by 10,000 this is

$$\begin{Bmatrix} 63{\cdot}3129 & -2{\cdot}4629 \\ -2{\cdot}4629 & 32{\cdot}7340 \end{Bmatrix},$$

giving the adjustments

$$\delta p_1 \quad 0{\cdot}264 \text{ per cent.}, \qquad \delta p_2 \quad 0{\cdot}162 \text{ per cent.},$$

and the corrected estimates

$$p_1 \quad 33{\cdot}597 \text{ per cent.}, \qquad p_2 \quad 70{\cdot}162 \text{ per cent.},$$

with standard errors

$$\pm 7{\cdot}957 \quad \text{and} \quad \pm 5{\cdot}721.$$

We may now partition the 442 parents observed, both phenotypically and genotypically as follows :—

	Homozygous	Heterozygous	Total
Strong	61·8	31·2	93
Medium	62·1	145·9	208
Weak	0	42·0	42
Total . . .	123·9	219·1	343

The agreement between the number of homozygotes estimated from the families without reference to negative parents, namely

123·9, with 122·6 the number estimated from the number of negative parents without reference to the families, is quite striking.

TABLE 3

Scores for each class of family, $p_1 = 0{\cdot}3$, $p_2 = 0{\cdot}7$

(*a*) Scoring for p_1

	Families	Score p_1	$i_{p_1 p_1}$
Strong mated to Strong			
With Negative children	3	+18·0000	65·6313
Without Negative children: 1 child	2	−0·3429	0·0588
2 children	5	−1·5328	0·4699
3 ,,	1	−0·4119	0·1697
4 ,,	2	−0·9864	0·4865
6 ,,	1	−0·6031	0·3637
		+14·1229	67·1799
Strong mated to Negative			
With Negative children	5	+15·0000	45·0000
Without Negative children: 1 child	5	−3·0000	1·8000
2 children	3	−3·0000	3·0000
3 ,,	3	−3·7059	4·5779
4 ,,	2	−2·7273	3·7191
5 ,,	3	−4·2923	6·1413
		−1·7255	64·2383
Strong mated to Weak			
With Negative children	0	...	...
Without Negative children: 1 child	2	−0·5455	0·1488
2 children	3	−1·5366	0·7870
3 ,,	2	−1·4323	1·0257
		−3·5144	1·9615
Strong mated to Medium			
With Negative children	2	+6·0000	18·0000
Without Negative children: 1 child	8	−1·4867	0·2763
2 children	10	−3·4107	1·1633
3 ,,	8	−3·7423	1·7506
4 ,,	5	−2·8466	1·6206
5 ,,	1	−0·6495	0·4219
6 ,,	2	−1·4239	1·0137
13 ,,	1	−0·8850	0·7832
		−8·4447	25·0296
		$i_{p_1 p_2}$	11·9189
		$i_{p_2 p_2}$	5·6756
		Score p_2	−4·0213

5. DISCUSSION

It is clearly demonstrated in the foregoing sections that one important cause of variation in the strength of the *P* reaction lies simply in the difference between persons with two and persons with only one positive gene. Other genetical causes are possible, but it is not easy, without family studies freed from the quantitative effects of age, even to prove their existence, still less to exhibit their nature.

Strength of reaction is generally observed to be associated in parents and offspring. Nevertheless, the view that all causes other than homozygosity were exogenic could not be disproved without a quantitative study showing that the degree of association observed is really greater than that to be expected merely from the association of homozygosity in parents and children.

TABLE 3 (*continued*)

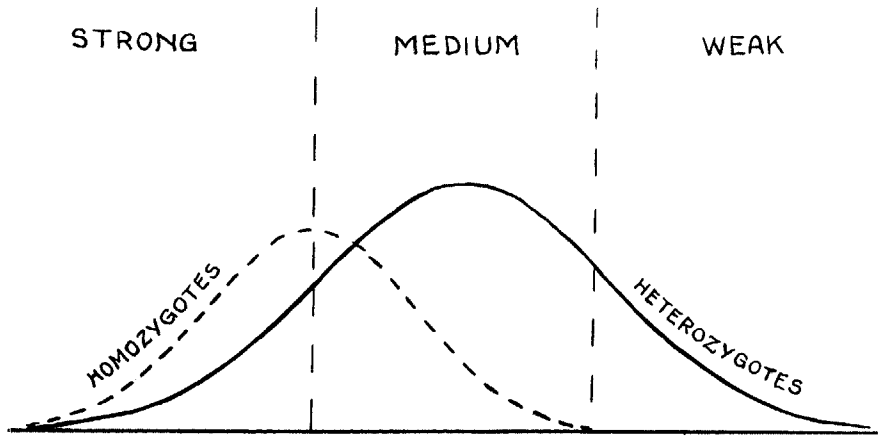
(*b*) Scoring for p_2

	Families	Score p_2	$i_{p_1 p_2}$
Strong mated to Medium	...	−4·0213	5·6756
Medium mated to Medium			
With Negative children	16	+45·7143	117·9046
Without Negative children: 1 child	9	−3·5897	1·4318
2 children	9	−7·0169	5·4705
3 ,,	10	−11·2928	12·7527
4 ,,	4	−5·7562	8·2835
5 ,,	3	−5·1148	8·7204
6 ,,	1	−1·9270	3·7133
7 ,,	1	−2·1217	4·5016
		+8·8954	162·7784
Medium mated to Weak			
With Negative children	6	+8·5714	12·2449
Without Negative children: 1 child	2	−0·6061	0·1837
2 children	6	−3·7838	2·3862
3 ,,	3	−2·9134	2·8293
4 ,,	3	−3·9326	5·1551
6 ,,	2	−3·8721	7·4966
		−6·5366	30·2958
Medium mated to Negative			
With Negative children	23	+32·8571	46·9387
Without Negative children: 1 child	8	−6·1538	4·7337
2 children	8	−12·6316	19·9447
3 ,,	1	−2·2581	5·0990
5 ,,	1	−3·0097	9·0583
8 ,,	1	−3·2903	10·8261
10 ,,	1	−3·3225	11·0390
		+2·1911	107·6395

Without being wholly exogenic the remaining variation in strength of reaction might be due partly to genes of small effect acting as cumulative factors, or to multiple alleles at the *P* locus. In the latter case there might be, for example, only two such positive alleles, or equally there might be very many with various grades of strength. Whichever of these views may be favoured, the existence of considerable non-genetic influence must be admitted, and this consideration should prevent any tendency to identify arbitrary, though striking, phenotypic grades with genetic entities.

At the present stage I suggest that we can express our knowledge of the variability of reactive strength in heterozygotes and homozygotes in the accompanying figure, in which the homozygotes are given just one half of the variance of the heterozygotes, as would be the case if there were a number of allelic positive genes interacting without dominance in respect of strength of reaction, and in such a way that the difference in titre between two homozygotes was the same as that between the corresponding heterozygotes. Whatever may be the interpretation, however, the diagram will serve to illustrate

DISTRIBUTION INTO PHENOTYPIC CLASSES
OF HOMOZYGOTES (PP)
AND HETEROZYGOTES (Pp)



the variability of the two genotypes in relation to the boundaries dividing Strong from Medium, and Medium from Weak, the considerable overlap between their ranges, and the improbability, in view of this variability, of any distinct bimodality or multimodality being recognisable.

An important fact antagonistic to the existence of multiple alleles is the failure of absorption to indicate any qualitative difference between Strong and Weak reactors. This is in contrast to such cases as A_1 and A_2, or to the quasi-alleles of the Rhesus system.

6. SUMMARY

An analysis of Henningsen's family data, in which parents are classified as Strong, Medium and Weak reactors to anti-*P* serum shows that these three phenotypic classes contain very unequal proportions of heterozygotes.

Subject to errors of random sampling, the proportions are estimated to be

Strong	. . .	33·597 per cent. heterozygous
Medium	. . .	70·162 ,, ,,
Weak	. . .	100·000 ,, ,,

The total number of homozygotes estimated in this way, 123·9, agrees closely with the number 122·6, estimated from the proportion of Negative parents.

Of the total number enumerated it appears that the homozygotes are about equally divided between the Strong and Medium phenotypes; while of the heterozygotes about 14 per cent. are Strong, 67 per cent. Medium, and 19 per cent. Weak (see figure).

Homozygosity is therefore a well-established cause of the variation of reactive strength. As to the residual causes, it would appear premature yet to infer even that these are of genetic origin, still less that they can be ascribed to a series of recognisable alleles at the *P* locus.

7. REFERENCES

LANDSTEINER, K., AND LEVINE, P. 1927. Further observations on individual differences of human blood. *Proc. Soc. Exp. Biol. N.Y.*, *24*, 941-942.

HENNINGSEN, K. 1949. Investigations on the blood factor *P*. *Acta Path. Scand.*, *26*, 639-654.

RACE, R. R., AND SANGER, R. 1950. *Blood Groups in Man*. Blackwell, Oxford.