THE GENETIC ANALYSIS OF FAMILIAL TRAITS.

I. SINGLE GENE SUBSTITUTIONS.

By LANCELOT HOGBEN.

(*From the Department of Social Biology in the University of London.*)

I. Introduction.

The smallness of the average human family introduces difficulties in testing Mendelian hypotheses already subjected to considerable discussion by Lenz (1929), Bernstein (1929–30), Weinberg (1927) and Dahlberg (1930)[1]. Families in which recessives occur will be families produced by parents who are both heterozygous and therefore indistinguishable from normal individuals, or by parents, one or both of whom are recessive. In a system of random mating the proportion of recessives having parents both of whom are normal will be $(1 - r)^2$, where r^2 is the frequency of recessives in the general population. Thus if the recessive character is a rare one, recessives will practically never have parents who exhibit the recessive character. The condition will exhibit itself as a "familial" rather than as a "hereditary" trait in the sense in which these terms are used by pathologists. When a single gene is involved there will be three families of single children without the recessive trait for every family of one recessive who is a single child. If all families containing recessives are taken from a human population the ratio of the total number of recessives to the total number of individuals in such a group of families will thus be greatly in excess of the theoretical proportion of recessives in all families of two heterozygous parents.

If p is the probability that the offspring of two heterozygous parents will be recessive ($p = \frac{1}{4}$, $\frac{1}{16}$, etc.) and $(p + q) = 1$, the probability that all the members of a family of s individuals will be normal is q^s, and the probability that such a fraternity will contain at least one recessive is $(1 - q^s)$. Corresponding to the total number of recessives r_s in n_s families of s individuals with at least one recessive there should be a theoretical population of normal individuals of $s.n_s + s.n_s.\frac{q^s}{1 - q^s} = \frac{s.n_s}{1 - q^s}$. If c

[1] Vide Sjogren, *Hereditas*, XIV, 1930.

is the maximum size of family, a familial condition may be inferred to be a recessive condition, if

$$\frac{\sum_{s=1}^{s=c} r_s}{\sum_{s=1}^{s=c} \frac{s.n_s}{1-q^s}} = p.$$

This is the formula proposed by Lenz. To subject the hypothesis that a familial trait is determined by one or more recessive genes we have to decide how much the quantity

$$\frac{\Sigma r_s}{\Sigma \frac{s.n_s}{1-q^s}},$$

may differ from p through errors of sampling. The object of this communication is to test the significance of extant data by a simple formula for calculating the standard deviation of p determined in this way.

Families of at least one recessive will be made up of 1, 2, 3, ... s recessives. The error involved in determining the ratio of recessives to normal offspring of heterozygous parents may be regarded as arising out of errors in the distribution of 0, 1, 2, ... s recessives in a given sample of s-membered families, the sample being taken so as to exclude the case when the number of recessives is 0. The problem therefore resolves itself into ascertaining to what error the value of p is subject when the method of sampling is subject to this restriction.

Let us suppose that we knew all the families of s individuals whose parents were heterozygotes, the expected number of recessives in a family of s would then be given by the binomial series:

Size of family (s)	Number of recessives							
	0	1	2	3	...	r	...	s
1	q	p						
2	q^2	$2qp$	p^2					
3	q^3	$3q^2p$	$3qp^2$	p^3				
r	q^r	$rq^{r-1}p$	.....			p^r		
s	q^s	$sq^{s-1}p$	.....			${}^sC_r q^{s-r}p^r$	...	p^s

Now the probability that a family of s individuals will contain r recessives is ${}^sC_r p^r q^{s-r} = P_r$; and if there are N_s families of s individuals, the number of families of s individuals which have r recessives is $N_s.P_r$, so that the number of recessives in all families of s individuals of which r are recessive is $r.N_s.P_r$. The total number of recessives in the population of families containing s individuals is therefore

$$\sum_{r=1}^{r=s} r.N_s.P_r \qquad \text{.....(1)},$$

or

$$N_s(1P_1 + 2P_2 + 3P_3 + \ldots + sP_s) = Q_s \qquad \text{.....(2)},$$

and the total population of s-fold families is $s.N_s$, so that the expectation of recessive offspring of heterozygous parents of s children is

$$\frac{N_s(1P_1 + 2P_2 + 3P_3 + \ldots + sP_s)}{sN_s} \qquad \ldots\ldots(3).$$

For the entire population s may have any value from 1 to c the maximum size of family, so that the expectation of recessives in a population of all families of whatever size is

$$\frac{\sum_{s=1}^{s=c}\sum_{r=1}^{r=s} r.N_s.P_r}{\sum_{s=1}^{s=c} s.N_s} = \frac{\sum_{s=1}^{s=c} Q_s}{\sum_{s=1}^{s=c} s.N_s} = p \qquad \ldots\ldots(4).$$

Let σ_R be the standard deviation of p, and let σ_s be the standard deviation of Q_s, *i.e.* of the quantity

$$N_s.(P_1 + 2P_2 + 3P_3 + \ldots + sP_s).$$

Then from elementary principles

$$\left[\sum_{s=1}^{s=c} s.N_s\right]^2.\sigma_R{}^2 = \sigma_1{}^2 + \sigma_2{}^2 + \sigma_3{}^2 + \ldots + \sigma_s{}^2 \qquad \ldots\ldots(5).$$

It remains only to determine the form of the terms $\sigma_1, \ldots \sigma_s$, etc.

Professor Bowley has suggested to me that this can be done by using a theorem of Sheppard's, for which a preliminary lemma is required. This theorem is as follows. If a universe contains N measurable objects y_1 being of measurement x_1, y_2 of measurement x_2, ... y_z of measurement x_z, so that $y_1 + y_2 + y_3 + \ldots + y_z = 1$, then if F is a linear function of y_1, y_2, etc. such that

$$F = (a_1y_1 + a_2y_2 + a_3y_3 + \ldots + a_ry_r + \ldots + a_zy_z) \qquad \ldots\ldots(6),$$

the standard deviation σ_F of this function is given by the equation

$$\sigma_F{}^2 = \frac{1}{n}\{\Sigma\, a_r{}^2y_r - F^2\} \qquad \ldots\ldots(7),$$

where n successive samples are taken to determine the value of F.

This follows from the following considerations:

Suppose that a sample of p objects be taken from N, the proportions at measurements $x_1, x_2, \ldots x_z$ will be $(y_1+\delta y_1)$, $(y_2+\delta y_2)$, ... $(y_z+\delta y_z)$, etc., so that

$$\Sigma\delta y_r = 0.$$

Let D be the sum of all the residuals except δy_r and δy_s. Then $D+\delta y_r+\delta y_s=0$, and similarly let

$$Y = 1 - y_r - y_s.$$

Now the standard deviation of y_t is

$$\sqrt{\frac{\Sigma\delta y_t^2}{n}}=\sqrt{\frac{y_t(1-y_t)}{n}} \qquad \ldots\ldots(8),$$

where n successive observations are made upon it. And we may write

$$D^2-\delta y_r^2-\delta y_s^2=2\delta y_r\delta y_s \qquad \ldots\ldots(9).$$

$$\therefore \frac{2\Sigma\delta y_r\delta y_s}{n}=\frac{\Sigma D^2}{n}-\frac{\Sigma\delta y_r^2}{n}-\frac{\Sigma\delta y_s^2}{n}$$

$$=\frac{1}{n}[Y(1-Y)-y_r(1-y_r)-y_s(1-y_s)]$$

$$=-\frac{2y_ry_s}{n}.$$

$$\therefore \Sigma\delta y_r\delta y_s=-y_ry_s.$$

Now $$\delta F=a_1\delta y_1+a_2\delta y_2+\ldots+a_z\delta y_z$$

and $$\delta F^2=a_1^2\delta y_1^2+a_2^2\delta y_2^2+\ldots+2a_1a_2\delta y_1\delta y_2,\ \ldots \text{ etc.}$$

$$\therefore \sigma_F^2=\Sigma\frac{\delta F^2}{n}=a_1^2\Sigma\frac{\delta y_1^2}{n}+a_2^2\Sigma\frac{\delta y_2^2}{n}+\ldots+2a_1a_2\Sigma\frac{\delta y_1\delta y_2}{n}+\ldots+2a_ra_s\Sigma\frac{\delta y_r\delta y_s}{n}$$

$$=\frac{1}{n}\{\Sigma a_t^2y_t(1-y_t)-2\Sigma a_ra_sy_ry_s\}$$

$$=\frac{1}{n}\{\Sigma a_t^2y_t-F^2\}.$$

Returning to our main theme, the quantity Q_s may be written

$$N_s(1-q^s)\times\left\{1\frac{P_1}{1-q^s}+2.\frac{P_2}{1-q^s}+3\frac{P_3}{1-q^s}+\ldots+s.\frac{P_s}{1-q^s}\right\} \qquad \ldots\ldots(10),$$

so that $$\frac{P_1}{1-q^s}+\frac{P_2}{1-q^s}+\ldots=1,$$

and the quantity within the brackets in (10) is of the same form as F in (6), the constants $a_1a_2a_3, \ldots a_z$ being replaced by 1, 2, 3, ... s.

Now n_s families are available for determining the value of this function. So we may write

$$\sigma_s^2=N_s^2(1-q^s)^2.\frac{1}{n_s}\left\{\frac{1}{1-q^s}(P_1+2^2P_2+3^2P_3+\ldots)\right.$$

$$\left.-\frac{1}{(1-q^s)^2}(P_1+2.P_2+3P_3+\ldots+s.P_s)^2\right\} \quad \ldots(11),$$

$$\sigma_s^2=N_s\left\{\sum_{r=1}^{r=s}r^2{}^sC_rp^rq^{s-r}-\frac{1}{1-q^s}\sum_{r=1}^{r=s}(r.{}^sC_rp^rq^{s-r})^2\right\} \quad \ldots(12).$$

The reduction may be effected as follows:

$$\frac{\partial}{\partial p}\left\{\sum_{r=1}^{r=s} {}^sC_r p^r q^{s-r}\right\} = \frac{\partial}{\partial p}\cdot\{(p+q)^s - q^s\},$$

$$\therefore \Sigma r\,{}^sC_r p^{r-1} q^{s-r} = s\,(p+q)^{s-1},$$

$$\therefore \Sigma r\,{}^sC_r p^r q^{s-r} = ps\,(p+q)^{s-1} \qquad \ldots\ldots(13),$$

$$\frac{\partial}{\partial p}(\Sigma r\,.\,{}^sC_r p^r q^{s-r}) = \frac{\partial}{\partial p}[ps\,(p+q)^{s-1}].$$

$$\therefore \Sigma r^2\,{}^sC_r p^{r-1} q^{s-r} = s\,(p+q)^{s-1} + ps\,(s-1)\,(p+q)^{s-2}.$$

$$\therefore \Sigma\, r^2\,{}^sC_r p^r q^{s-r} = ps\,(p+q)^{s-1} + p^2 s\,(s-1)\,(p+q)^{s-2} \quad \ldots(14).$$

Putting $(p+q) = 1$ and substituting the expressions (13) and (14) in (12) we have

$$\sigma_s{}^2 = N_s\left\{ps + p^2 s\,(s-1) - \frac{1}{1-q^s}\cdot p^2 s^2\right\}$$

$$= \frac{N_s\,(1-q).s}{1-q^s}\cdot\{sq^s\,(q-1) + q\,(1-q^s)\}.$$

By equation (5) we have

$$\sigma_R{}^2 = \frac{\sum_{s=1}^{s=c} N_s\cdot\frac{(1-q)\,s}{1-q^s}\cdot\{sq^s\,(q-1) + q\,(1-q^s)\}}{\left[\sum_{s=1}^{s=c}.s.N_s\right]^2}$$

$$= \frac{\sum_{s=1}^{s=c}\frac{s.n_s\,(1-q)}{(1-q^s)^2}\cdot\{sq^s\,(q-1) + q\,(1-q^s)\}}{\left[\sum_{s=1}^{s=c}\cdot\frac{s.n_s}{1-q^s}\right]^s}.$$

Now putting $$c_s = \frac{s}{1-q^s}$$

and $$k_s = \frac{s\,(1-q)}{(1-q^s)^2}\cdot\{sq^s\,(q-1) + q\,(1-q^s)\},$$

we may write $$\sigma_R = \frac{\sqrt{\sum_{s=1}^{s=c} n_s.k_s}}{\sum_{s=1}^{s=c} n_s.c_s},$$

where σ_R is the standard deviation of the ratio p, and the values of c_s and k_s for $q = \frac{3}{4}$ from $s = 1$ to $s = 20$ are given in the accompanying tables.

It is easily seen that when $s=1$ the value of this expression becomes zero, since there is no error involved in the calculation based on the

foregoing method of adjustment. Further it is evident that the limit of sq^s as s approaches infinity is zero. Thus the above expression approximates to

$$\sqrt{\frac{q(1-q)}{s.n_s}}$$

when the size of the family is large. This is equivalent to the customary formula for the standard deviation of a ratio.

II. Recorded data of familial traits.

The formulae proposed in the preceding remarks will now be applied to test the significance of adjusted ratios obtained by applying the Lenz formula to recorded pedigrees of traits which occur in offspring of parents who do not themselves exhibit them. In three of the examples chosen the condition dealt with appears to be genetically heterogeneous, and some reference must be made to the distinction between separate classes of genetical phenomena included under the same diagnostic category.

(i) *Albinism.*

Over six hundred pedigrees of albinism have been collected by Pearson, Usher and Nettleship. In the ensuing table are excluded all non-European families, sibships in which partial albinism occurs, sibships in which the sex of some of the sibs is uncertain, sibships of which either parent is an albino, and sibships where reference to the explanatory notes indicates that albinism may have been incomplete[1].

[1] *Albinism* (pedigree numbers in Pearson, Nettleship and Usher):

One: 14, 17, 30, 41, 61, 64, 93, 152, 121, 169, 194, 197, 239, 250, 318, 453, 460, 464, 530, 533, 531, 561, 571, 649.

Two: 32, 33, 70, 83, 100, 151, 161, 211, 223, 223, 225, 245, 249, 305, 383, 450, 463, 467, 478 *b*, 550, 555, 560.

Three: 3, 20, 27, 30, 103, 108, 113, 135, 137, 179, 186, 187, 190, 198, 225, 232, 240, 241, 296, 313, 381, 453, 455, 467, 472, 478 *a*, 540, 577, 640.

Four: 8, 31, 49, 96, 132, 181, 192, 207, 319, 376, 384, 396, 453, 468, 471, 493, 493, 534, 548, 559, 560.

Five: 36, 37, 71, 85, 86, 86, 98, 118, 131, 146, 185, 222, 258, 265, 301, 311, 366, 397, 458, 462, 465, 498, 596, 612, 642.

Six: 30, 31, 69, 193, 210, 220, 233, 256, 258, 303, 261, 314, 320, 365, 374, 383, 405, 457, 460, 480, 495, 538, 568, 578, 634, 640, 644.

Seven: 2, 28, 150, 193, 211, 211, 221, 204, 227, 412, 450, 459, 460, 465, 596, 557, 652, 654.

Eight: 6, 112, 112, 119, 179, 183, 218, 308, 412, 459, 479, 484.

Nine: 10, 28, 99, 116, 177, 219, 382, 482.

Ten: 43, 90, 453, 474, 539, 635, 637, 646.

Eleven: 27, 95, 100, 298, 460, 546.

Twelve: 58, 136, 313, 365.

Thirteen: 59.

For the entire series of sibships the corrected value of p is 0·286. The theoretical value of 0·250, determined in this way, has a standard deviation of 0·0083. The error is approximately 4·5 times the standard deviation.

The percentage of first cousin marriages in this series is 17. There is an unexplained excess of males to females in albinos of 128 : 100, but there is also an excess of males in the ratio 118 : 100 in the entire group. None of the pedigrees provide clear evidence of Nasse's Law. If there were two forms of albinism, one recessive and autosomal and one recessive and sex-linked, we should expect an excess of females in the normal sibs. It seems possible that the gene albinism is semi-lethal in the female but not in the male.

Size of fraternity	Number of fraternities	Albinos	$n_s c_s$	Corrected value of p	$n_s k_s$
1	24	24	96·0	—	00·0000
2	22	27	100·6	0·269	2·6939
3	29	36	150·5	0·239	7·6261
4	21	38	122·9	0·309	8·8211
5	25	49	163·9	0·299	14·7944
6	27	66	197·1	0·299	20·9506
7	18	42	145·4	0·289	17·4643
8	12	35	105·7	0·331	14·0684
9	8	23	77·8	0·297	11·0413
10	8	16	84·7	0·189	12·7334
11	6	22	68·9	0·314	10·8316
12	4	13	49·6	0·263	8·0784
13	1	3	13·3	0·225	2·2335

(ii) *Retinitis pigmentosa.*

The recorded cases of retinitis pigmentosa fall into two groups. In the Nettleship Memorial volume of the *Treasury of Human Inheritance*, Dr Julia Bell has collected 250 recorded pedigrees, to which have since been added two long pedigrees of Usher and Shennan. In the first group are included the last-named pedigrees, together with four others, two recorded by Nettleship, one by Schneider and one by Snell. In this group matings of two normal individuals never give affected individuals. The trait behaves as an "hereditary" disorder; but the proportion of pathotypes is less than would be expected for a completely dominant mutant condition.

In the remainder two normal parents give affected offspring and there is a high incidence of consanguineous unions. Many of these pedigrees are short and the number of affected individuals is greatly in excess of the theoretical expectation for a recessive mutant when corrected by the method adopted in the foregoing example. This may be attributed

(1) to the likelihood that fraternities with a high familial incidence will tend to be placed on record more frequently than isolated cases, (2) to the possibility that this group contains a considerable proportion of an incompletely dominant type. In the Nettleship Memorial volume forty-one extensive pedigrees investigated by Usher are not open to either of these objections, and these have been summarised in the accompanying table. Stillborn individuals and miscarriages are not counted in assigning the size of fraternity. There are ninety affected individuals in fifty-three sibships. Of these fifty-five are males and thirty-five females. As in the case of albinism, there is an unexplained excess of males, but none of the pedigrees give evidence of Nasse's Law. In 17 per cent. of the sibships the parents are first cousins. The corrected value of p is 0·216 and the standard deviation of the theoretical value 0·25, calculated by the method outlined, is $\pm$ 0·017, so that the error is twice its standard deviation.

Usher Pedigrees[1].

Size of fraternity	No. of fraternities	Pathotypes	$n_s c_s$	$n_s k_s$
1	4	4	16·0	0·0000
2	2	2	9·14	0·2449
3	3	3	15·57	0·7889
4	8	12	46·81	3·3604
5	5	9	32·77	2·9590
6	3	4	21·90	2·3278
7	7	16	56·55	6·7197
8	5	10	44·45	5·8618
9	2	4	19·46	2·7603
10	1	1	10·59	1·5917
11	5	14	57·42	9·0263
12	4	8	49·57	8·0782
13	1	1	13·32	2·2335
14	2	3	28·51	4·8927

[1] Usher pedigrees:

One: 142, 153, 156, 169.
Two: 147, 157.
Three: 138, 154, 164.
Four: 136, 137, 138, 144, 151, 156, 162, 164.
Five: 134, 136, 152, 165, 170.
Six: 141, 143, 163.
Seven: 133, 140, 149, 150, 155, 159, 166.
Eight: 139, 156, 168, 171, 172.
Nine: 148, 158.
Ten: 134.
Eleven: 136, 139, 144, 145, 167.
Twelve: 135, 133, 147, 161.
Thirteen: 160.
Fourteen: 146, 173.

(iii) *Haemophilia.*

The next three examples conform to Nasse's Law. Algebraically the problem is identical with that of an autosomal recessive condition. The proportion of affected male offspring of normal parents should be a quarter of the entire progeny if the condition is sex-linked. To minimise the danger of prejudicing the result by including cases placed on record because a high familial incidence has attracted attention, the following table is based on pedigrees of haemophilia extending over at least five generations, as summarised by Bulloch and Fildes[1]. Only unions of two normal parents are considered.

Size of fraternity	Number of fraternities	No. of male bleeders	$n_s c_s$	$n_s k_s$
1	12	12	48·0	00·0000
2	25	35	114·3	3·0812
3	36	60	186·8	9·4668
4	29	52	169·7	12·1814
5	16	29	104·9	9·4684
6	23	51	167·9	17·8468
7	17	36	137·3	16·4941
8	10	27	88·9	11·7237
9	10	24	97·3	13·8016
10	11	27	116·6	17·5084
11	9	35	103·4	16·2473
12	5	19	62·0	10·0978
13	7	24	93·2	15·7346
14	2	8	28·5	4·8927
15	0	0	0	0·0000
16	3	14	48·5	8·6002
17	1	3	17·1	3·0738
18	0	0	0	0·0000
19	1	8	19·1	3·4814
20	1	4	20·1	3·6821
24	1	2	24	4·1662

This gives a value for p of 0·285 and the standard deviation of the theoretical value of 0·250 is $\pm$ 0·008. The error is about four and a half times the standard deviation.

[1] *Haemophilia* (pedigree numbers from Bulloch and Fildes):

One: 407, 408, 428, 431, 407, 409, 480, 481, 494.

Two: 373, 373, 373, 391, 407, 407, 407, 378, 407, 408, 408, 408, 412, 417, 428, 474, 475, 481, 497, 412, 464, 481, 481, 494, 494.

Three: 373, 373, 373, 388, 388, 407, 377, 408, 408, 408, 408, 412, 378, 427, 428, 431, 436, 441, 380, 455, 460, 460, 462, 468, 468, 474, 480, 480, 497, 464, 464, 581, 594, 604, 604.

Four: 373, 373, 373, 373, 391, 394, 378, 378, 401, 401, 407, 407, 408, 409, 412, 417, 426, 431, 431, 436, 436, 494, 408, 408, 464, 464, 481, 481, 494.

Five: 380, 389, 394, 398, 401, 401, 407, 409, 417, 417, 428, 455, 462, 468, 508, 604.

Six: 373, 373, 388, 389, 389, 389, 389, 390, 391, 391, 398, 407, 407, 417, 426, 428, 455, 475, 493, 561, 594, 594, 604.

Seven: 373, 408, 418, 436, 440, 455, 389, 389, 468, 480, 497, 509, 594, 604, 604, 604, 606.

(iv) *Congenital stationary night blindness.*

All the cases of congenital stationary night blindness set forth in Dr Julia Bell's monograph may be classified in four groups:

(1) The first is represented by a single pedigree over six generations, that of Sinclair (300). No cases of matings between normal individuals giving affected individuals occur in this tree. In six sibships with one affected parent, there are twelve affected individuals out of a total of twenty-six. Of these twelve five are females. Qualitatively and quantitatively the data derived from this source are compatible with the view that one form of congenital stationary night blindness is determined by a single completely dominant mutant gene. To this may be added the Sedan pedigree (301). No affected individuals appear among the offspring of normal parents; but there is a significant excess of affected individuals in families containing both types. In eight sibships of twenty-eight individuals in all there are nineteen affected individuals of which ten are female. On the hypothesis which fits the Sinclair pedigree the number of affected individuals should be $14 \pm 2{\cdot}55$. The odds against this result are not very great. In both the Sedan and Sinclair pedigrees the pathotypes are in excess. Taking them together the theoretical expectation is $27 \pm 3{\cdot}7$; and the observed number of pathotypes is 31. The discrepancy is not significant.

(2) A second group is represented by the celebrated Cunier tree. In this family, which has been worked out over ten generations, there occur 131 affected individuals in sibships of which one parent is normal. There is no significant difference between the number of affected males and females. No affected individuals have normal parents, but in every generation except the sixth the number of normal individuals in families which contain affected individuals is greatly in excess of the number required by the hypothesis that this type of night blindness is determined

Eight: 373, 388, 407, 417, 431, 468, 377, 391, 475, 606.
Nine: 378, 378, 378, 417, 436, 440, 474, 497, 508, 512.
Ten: 388, 389, 389, 418, 426, 426, 427, 461, 481, 561, 581.
Eleven: 373, 389, 427, 468, 468, 378, 475, 497, 603.
Twelve: 380, 390, 475, 581, 606.
Thirteen: 373, 389, 401, 377, 433, 497.
Fourteen: 468, 512.
Fifteen: —
Sixteen: 481, 581, 606.
Seventeen: 391.
Eighteen: —
Nineteen: 389.
Twenty: 581.
Twenty-four: 379.

by a single completely dominant gene. Altogether there are 231 normal and 131 affected individuals in families of which one parent is night blind. In this pedigree the condition is completely dominant, since it never skips a generation. It may be regarded as belonging to the foregoing group, if we assume linkage with a lethal or semi-lethal gene. A high incidence of cousin marriages supports this supposition.

(3) A third, the largest, group of cases, matings of normal parents, may give affected males but never affected females. The affected males are greatly in excess. The only exception is in the sixth generation of the Nettleship and Stanford Morton pedigree, in which one unaccountable female occurs. The ensuing table is a summary of three pedigrees by Nettleship (327, 319, 323), one by Newman (321), Nettleship and Stanford Morton (325), Pagenstecher (318), Cant (312), Cutler (325), Pfluger (326) and Amman (332).

Size of family	Number of families	$n_s c_s$	Number of affected males	$n_s k_s$
1	2	8·00	2	0·0000
2	5	22·86	6	0·6122
3	6	31·13	9	1·5778
4	9	52·66	14	3·7804
5	5	32·75	10	2·9589
6	3	21·90	5	2·3278
7	8	64·62	23	7·7619
8	3	26·67	8	3·5171
9	2	19·46	5	2·7603
10	2	21·19	6	3·1833
11	2	22·97	4	3·6105
12	—	—	—	0·0000
13	1	13·31	2	2·2335

This gives $p = 0{\cdot}278$ and the calculated standard deviation of the theoretical value 0·25 is $\pm$ 0·017. The error is less than twice the standard deviation.

(4) A fourth group includes sibships of which both parents are normal and no excess of males is found among the affected individuals, affected females occurring among offspring of normal parents. These pedigrees are those of Cutler (325, 306), Fuchs (303, 307, 313), Fitzgerald (310), Vieusse (315), Swanzy (316), Nettleship (312). They are summarised in the following table.

Size of fraternity	Number of fraternities	Number of pathotypes	$n_s c_s$	$n_s k_s$
3	3	7	15·57	0·7889
4	1	1	5·85	0·4200
6	1	2	7·30	0·7759
8	1	3	8·80	1·1724
10	1	5	10·59	1·5917
13	1	3	13·31	2·2335
14	1	2	14·25	2·4463

This gives as the corrected proportion of pathotypes 0·304. The standard deviation of the theoretical value 0·25 is $\pm$ 0·040, on the assumption that this group is determined by a single recessive autosomal gene. There are no data regarding consanguinity to support this conclusion.

(v) *Colour blindness*[1].

With the exception of the Cunier tree which shows matrilineal transmission the recorded pedigrees of simple congenital red-green colour blindness are qualitatively consistent with the hypothesis that the condition differs from the normal on account of a single recessive gene substitution located on the *X*-chromosome. The cases where supposedly normal fathers have colour-blind daughters are so rare that they may legitimately be attributed to uncertain paternity, uncertain diagnosis or mutation. Sibships containing colour-blind males with two normal parents have been extracted from the pedigrees compiled by Dr Julia Bell (Part II of the Nettleship Memorial Volume), excluding pedigrees extending over less than four generations. Sibships in which doubt exists concerning the sex or diagnosis of one or more the members have also been rejected in the table which ensues. In the whole group of over two hundred sibships the unaccountable occurrence of colour-blind females is only seen in ten.

[1] Colour blindness. The pedigree numbers of fraternities of one, two, etc., sibs in Dr Bell's memoir are as follows:

One: 407, 409, 433, 440, 449, 452, 582.
Two: 409, 412, 412, 412, 413, 415, 415, 419, 420, 422, 425, 427, 427, 432, 439, 446, 514, 580, 580, 581, 589.
Three: 406, 407, 408, 408, 409, 412, 412, 416, 419, 425, 427, 427, 432, 440, 440, 440, 441, 441, 441, 443, 445, 449, 450, 495, 495, 578, 582, 582, 582.
Four: 408, 408, 409, 409, 410, 411, 414, 415, 417, 419, 420, 420, 420, 424, 440, 440, 441, 441, 443, 443, 449, 450, 580, 580, 580, 581, 582, 582, 582, 587, 587, 589, 589, 589.
Five: 406, 407, 407, 407, 407, 408, 408, 408, 409, 412, 413, 415, 417, 419, 419, 424, 425, 425, 425, 428, 428, 439, 440, 441, 448, 450, 451, 452, 514, 514, 580, 580, 584, 587, 590.
Six: 406, 407, 407, 408, 408, 408, 409, 411, 417, 419, 422, 432, 441, 449, 580, 580, 580, 587.
Seven: 406, 407, 407, 408, 408, 412, 412, 415, 428, 494, 514, 588.
Eight: 409, 411, 417, 417, 428, 430, 440, 442, 450, 494, 494, 582.
Nine: 407, 407, 408, 409, 409, 418, 420, 423, 432, 443, 444, 449, 494, 587.
Ten: 409, 411, 411, 440, 443, 514, 582, 587.
Eleven: 409.
Twelve: 409.
Thirteen: 412, 440, 514.
Fourteen: 407.

Size of fraternity	Number of fraternities	Number of colour-blind males	$n_s c_s$	$n_s k_s$
1	8	8	32·00	0·000
2	21	26	95·97	2·564
3	29	38	150·51	7·627
4	34	50	198·90	14·348
5	35	59	229·60	20·720
6	18	24	131·40	13·968
7	12	28	96·96	11·742
8	12	29	106·68	14·064
9	14	32	136·22	19·320
10	9	20	95·31	14·310
11	1	4	11·49	1·805
12	1	3	12·40	2·019
13	3	13	39·93	6·699
14	1	4	14·25	2·446

In this group of pedigrees the proportion of colour-blind males in sibships with normal parents adjusted as before is 0·250. The theoretical proportion of 0·250 is subject to a standard deviation of $\pm$ 0·008. There is therefore a completely satisfactory agreement between hypothesis and observation. It is to be noticed that colour blindness is not a rare condition. There is therefore less likelihood that the recorded observations will be biassed by the endeavour to demonstrate a higher familial incidence of the condition than might be expected, if, like haemophilia, it were a rare condition.

(vi) *Juvenile amaurotic idiocy.*

The last example which will be taken is from Sjogren's recent memoir on juvenile amaurotic idiocy (*Hereditas*, XIV, 1930). The entire data are summarised in Table 4 of his paper. In fifty-nine fraternities 115 juvenile amaurotics occurred. Of these 61 were males and 54 females. The excess of males is not significant. The size and frequency of the fraternities is given below.

Size of fraternity	Number of fraternities	$n_s c_s$	Number of affected individuals	$n_s k_s$
1	7	28·0	7	0·0000
2	3	13·7	5	0·367
3	7	36·3	10	1·841
4	10	58·5	19	4·200
5	7	45·9	11	4·142
6	6	43·8	9	4·656
7	7	56·5	13	6·792
8	7	62·2	25	8·204
9	2	19·4	6	2·760
10	0	0·0	0	0·000
11	1	11·5	3	1·805
12	1	12·4	3	2·019
13	1	13·3	4	2·233

The adjusted value of p obtained by the foregoing method is 0·286,

and the standard deviation of the theoretical value of 0·250 is ± 0·016, the error being less than two and a half times the standard deviation.

(vii) *Leber's Disease (hereditary optic atrophy)*[1].

The recorded cases of Leber's Disease compiled by Dr Julia Bell fall into two groups. In the European pedigrees, the occurrence of females affected in fraternities with parents both normal is so rare that it may be attributed to mutation, illegitimacy or incorrect diagnosis. In the Japanese cases the proportion of males and females among affected offspring of normal parents is not significantly different. The following table is based upon the European pedigrees of more than two generations. Owing to the late onset of the disease sibships of which normal members died before attaining adulthood as well as sibships containing individuals of doubtful sex or diagnosis have been excluded.

Size of fraternity	Number of fraternities	Number of affected males	$n_s c_s$	$n_s k_s$
1	20	20	80·00	0·000
2	29	38	132·57	3·551
3	22	35	114·16	5·785
4	22	38	128·73	9·241
5	29	51	190·11	17·162
6	13	33	94·89	10·087
7	14	31	113·10	13·583
8	13	29	115·57	15·241
9	3	5	29·19	4·140
10	3	12	31·79	4·775
11	3	5	34·45	5·416
12	1	1	12·39	2·020
13	1	1	13·32	2·233

[1] Leber's disease. The pedigree numbers of the European cases on which this table is based are as follows, classified by size of fraternity:

One: 704, 707, 715, 716, 716, 716, 724, 743, 744, 745, 745, 748, 762, 762, 769, 790, 790, 790, 792, 837.

Two: 707, 716, 716, 728, 732, 733, 734, 734, 737, 738, 738, 740, 743, 765, 767, 767, 777, 777, 788, 791, 796, 830, 837, 837, 837, 837, 843, 885, 886.

Three: 704, 707, 707, 711, 712, 716, 719, 727, 733, 735, 737, 742, 743, 767, 778, 788, 788, 837, 838, 840, 843, 886.

Four: 704, 704, 707, 716, 716, 716, 716, 724, 724, 728, 733, 734, 741, 762, 788, 796, 837, 838, 842, 842, 910, 910.

Five: 705, 705, 707, 707, 710, 710, 716, 724, 728, 732, 737, 740, 743, 747, 748, 752, 768, 777, 781, 796, 830, 836, 837, 838, 910, 910, 910, 910.

Six: 707, 711, 716, 716, 728, 729, 738, 739, 740, 740, 761, 798, 827.

Seven: 704, 705, 707, 707, 710, 714, 728, 738, 739, 765, 768, 780, 832, 837.

Eight: 704, 707, 719, 724, 724, 724, 728, 737, 739, 756, 784, 795, 838.

Nine: 704, 731, 736.

Ten: 737, 778, 837.

Eleven: 778, 837, 837.

Twelve: 798.

Thirteen: 837.

As compared with the expected proportion of 0·25 the adjusted proportion of affected males in fraternities of two normal parents is 0·274. The standard deviation is 0·009. The discrepancy is less than three times the standard deviation.

The Japanese cases are qualitatively consistent with the hypothesis that the condition is determined by an autosomal recessive gene substitution. As the data are scanty all sibships of two normal parents have been pooled in the ensuing table, unless doubt exists as to sex or condition or an individual dies young. In such cases the fraternities containing such individuals have been rejected. Affected individuals of both sexes are classified together. The adjusted proportion of affected males is 0·318. The standard deviation of the theoretical proportion of 0·25 is 0·022. The discrepancy is more than three times the standard deviation.

Japanese cases.

Size of fraternity	Number of fraternities	Affected individuals	$n_s c_s$	$n_s k_s$
1	0	0	0	0·0
2	5	8	22·86	0·6122
3	3	3	15·57	0·7889
4	5	8	29·26	2·1003
5	2	4	13·11	1·1836
6	4	7	29·20	3·1038
7	6	13	48·47	5·8215
8	3	12	26·67	3·5171
9	2	10	19·46	2·7603

III. TABLE OF VALUES FOR c_s AND k_s ($q = \frac{3}{4}$).

S.	c_s	k_s
1	4	0
2	4·5714	0·12245
3	5·1892	0·26297
4	5·8514	0·42005
5	6·5557	0·59178
6	7·2991	0·77596
7	8·0783	0·97024
8	8·8900	1·1724
9	9·7306	1·3802
10	10·597	1·5917
11	11·485	1·8053
12	12·392	2·0196
13	13·316	2·2335
14	14·254	2·4464
15	15·203	2·6575
16	16·162	2·8667
17	17·129	3·0738
18	18·102	3·2787
19	19·081	3·4814
20	20·064	3·6821

IV. Summary.

1. If a familial trait is determined by a single recessive gene

$$\frac{R}{\Sigma \frac{s \cdot n_s}{1 - q^s}} = p,$$

where p is the probability that the offspring of two heterozygous parents will be recessive, R is the total number of recessives in observed families of at least one recessive, n_s is the number of observed s-membered fraternities containing at least one recessive, and $(p + q) = 1$.

2. The standard deviation of p determined in this way is

$$\frac{\sqrt{\Sigma n_s k_s}}{\Sigma n_s c_s},$$

where

$$c_s = \frac{s}{1 - q^s},$$

and

$$k_s = \frac{s\,(1 - q)}{(1 - q^s)^2} \cdot \{sq^s\,(q - 1) + q\,(1 - q^s)\}.$$

3. Tables of c_s and k_s are given for $s = 1$ to 20, q being $\frac{3}{4}$.

4. In general the proportion of recessives in recorded families showing either autosomal or sex-linked familial traits is significantly higher than $\frac{1}{4}$. This may be due to the fact that cases showing a high familial incidence are recorded in medical journals more frequently than cases showing a low familial incidence.

5. If this is so, little progress will be made in genetical analysis of pathological traits by compilation of recorded data. This conclusion is supported by the fact that two cases showing comparatively insignificant discrepancies were based on unweighted observations of individual investigators, and the only case of completely satisfactory agreement is a disorder which is not rare.