On the Selective Advantage of Cystic Fibrosis Heterozygotes

ALFRED G. KNUDSON, JR.,[1] * LOWELL WAYNE,[2] AND WILBUR Y. HALLETT[3] †

[1] *Department of Biology, City of Hope Medical Center, Duarte, California.*
[2] *Allan Hancock Foundation, University of Southern California, Los Angeles, California.*
[3] *Department of Medicine, City of Hope Medical Center, Duarte, California.*

THE HIGH FREQUENCY of the recessive gene for cystic fibrosis of the pancreas (CF) estimated for Caucasian populations is generally attributed to a selective advantage of the heterozygous carrier. The magnitude of this advantage would need to be only about 2% in order to maintain at equilibrium a gene frequency of approximately 0.02. In our own recent study (Hallett *et al.*, 1965), we included a survey for evidence of disease resistance among CF heterozygotes but were not able to find a difference between subjects and controls. We did notice a difference between the groups with respect to the average sizes of their sibships but were reluctant to draw attention to what could be a sampling error. We now feel that the problem should be considered further in view of the report of a similar finding from Australia (Danks *et al.*, 1965).

MATERIAL AND METHODS

The selection of subjects and procedure for study have been described (Hallett *et al.*, 1965). Although the purpose of the previous study was to establish whether heterozygous carriers might sustain some physiological disadvantage, inquiry into family histories provided the data necessary for the present report. The study group consisted of the parents of documented cases of CF, while the controls were for the most part friends and neighbors of the study group subjects. The close similarities of the two groups have been emphasized previously.

The information regarding families was obtained by a combination of questionnaires and interviews, in order to maximize accuracy. Data pertinent to the present study included information on the parents of the subjects

Received July 24, 1966.

Supported by U. S. Public Health Service research grant AM 05989 from the National Institute of Arthritis and Metabolic Diseases, National Institutes of Health.

* Present address: Medical Center, State University of New York at Stony Brook, Stony Brook, New York 11790.

† Present Address: University of Southern California School of Medicine, Los Angeles, California.

TABLE 1. COMPARISON OF NUMBERS OF LIVE OFFSPRING OF GRANDPARENTS

Number of offspring per family	CF families		Control families	
	Number	Total sibs	Number	Total sibs
1	9	9	15	15
2	20	40	37	74
3	33	99	23	69
4	15	60	15	60
5	21	105	11	55
6	10	60	7	42
7	10	70	2	14
8	6	48	2	16
9	—	—	3	27
10	3	30	2	20
11	1	11	—	—
12	2	24	—	—
13	1	13	—	—
14	—	—	1	14
TOTAL	131	569	118	406

Mean number of offspring	4.34		3.43
Variance	6.07		5.12
Variance of mean	0.0463		0.0434
Difference of means		0.91	
Sum of variances of means		0.0897	
Standard error of difference of means		0.30	
Difference of means / Standard error		3.0 ($P<0.01$)	

(Generation I), the subjects and their sibs (Generation II), and the children of the subjects (Generation III). The sib data included information on abortions, stillbirths, and postnatal deaths. All data were coded, recorded on punch cards, and originally examined by two of us with Dr. Frank Massey (Department of Biostatistics, School of Public Health, University of California at Los Angeles). The later scrutiny of data, which forms the basis for the present report, was performed by one of the authors (L. W.), using the facilities of the Computer Sciences Laboratory at the University of Southern California.

RESULTS

The numbers of live offspring of persons in Generation I are compared in Table 1. The mean numbers of 4.34 and 3.43 for CF and control families, respectively, are significantly different ($P<0.01$) by variance analysis.

These means are calculated for offspring at birth. We do have the additional observation that 19 of the CF offspring died during the first nine months of life, while only three of the control offspring died by then. The mean numbers of children surviving to this age are, therefore, 4.21 for CF and 3.41 for control. The difference of the means, 0.80, is still significant at the one per cent level. That this trend does not continue into adult life is shown by the observation that at the time of the study 506 of the offspring of the CF group and 356 of the offspring of the control group were still living; the means are 3.86 and 3.02, respectively, and the difference is 0.84.

We had already reported that in Generation III (eliminating of course those control families with no members in Generation III) the study families had more liveborn offspring (3.0 per family) than the control families (2.7 per family). Of course, this number could be influenced by the possibly greater willingness to participate on the part of CF families with more than one affected child, and so on the average more total children. We did wonder, however, whether the average family size in Generation II might be correlated with that in Generation III, so that the Generation II data might be biased. Evidence against this explanation is provided by the correlation coefficients for the two groups, $r = .049$ for CF, and $r = .209$ for controls.

DISCUSSION

In agreement with the Australian workers, we find that parents of children with CF come themselves from larger sibships than do control persons. The difference is an index of relative fertility, and the difference in survivors is an index of relative survival values. The magnitude of the relative disadvantage, s, of the normal homozygote compared with the heterozygote may be obtained from the following relationship:

$$\frac{1-s}{1} = \frac{\text{average sibship size for normals}}{\text{average sibship size for heterozygotes}} \qquad (1)$$

The values of s obtained from the Australian data and from our data are 0.10 and 0.21, respectively. Attention should be called to the fact that the incidence of childlessness in each group is assumed to be similar, although we have no way to check this assumption at present.

The difference is very large and would have an enormous effect on the frequency of this recessive lethal allele. At equilibrium, the relationship $q=s/(1+s)$ would apply, and even a value of $s=0.1$ would give a value of $q=0.09$ and a birth incidence of cystic fibrosis of about 0.8%. Obviously then, if this value of s is correct, the Caucasion population is not in equilibrium with respect to this gene. On the other hand, the observed disease incidence and presently estimated gene frequency could be attained if the advantage is of relatively recent origin and equilibrium not yet attained. Assuming that the coefficient of selection for CF homozygotes is 1.0, for heterozygotes 0.0, and for normal homozygotes s, then the gene frequency in a new generation (q') bears the following relationship to that in the preceding generation (q):

$$q' = \frac{q-q^2}{1-s(1-q)^2-q^2} = \frac{q}{1-s+(1+s)q} \qquad (2)$$

$$\Delta q = q' - q = \frac{sq - (1+s)q^2}{1 - s + (1+s)q} \tag{3}$$

The number of generations (n) required for the gene frequency to change from q_o to q_n is derived as follows: Setting $dq/dn = \Delta q$, then

$$dn = \frac{1 - s + (1+s)q}{sq - (1+s)q^2}\, dq$$

$$= \left(\frac{1-s}{s}\right)\frac{dq}{q} + \left(\frac{1+s}{s}\right)\left(\frac{dq}{s - (1+s)q}\right)$$

and

$$n = \int_0^n dn = \frac{1-s}{s} \ln(q_n/q_o) + \frac{1}{s} \ln\left(\frac{s - (1+s)q_o}{s - (1+s)q_n}\right) \tag{4}$$

If we estimate $q_n = 0.02$ and q_o to be approximately one-tenth of q_n, or 0.002, not an unreasonable estimate incidentally for the frequency among Oriental and Negro populations, then, for $s = 0.10$,

$$n = \frac{0.9}{0.1} \ln 10 + \frac{1}{0.1} \ln\left(\frac{0.1 - (1+0.1)0.002}{0.1 - (1+0.1)0.02}\right)$$

$$= 23 \text{ generations}$$

Therefore, a heterozygous carrier advantage which suddenly began operating 23 generations ago with a coefficient of selection of 0.1 could explain the presently observed CF incidence.

If this relative advantage of the carrier is real and still operating, then the gene frequency will continue to rise. The new frequency after one generation, given by equation (2), would, for $q = 0.02$ and $s = 0.1$, yield $q' = 0.022$. The new disease incidence would be q'^2, or 4.8 per 10,000, compared with the present estimate of 4.0 per 10,000. It will be observed that the *difference* (0.8 per 10,000) is of the order of magnitude of the incidence of such disorders as phenylketonuria and galactosemia.

The physiological basis for such an advantage of carriers would be of extreme interest of course. Of prime importance, however, is the necessity for establishing whether these observations are valid. Hopefully, other workers will report their findings.

SUMMARY

Grandparents of cystic fibrosis patients were found to have an average of 4.34 offspring compared to 3.43 for controls ($P<0.01$). Among survivors at the time of the study the difference was still significant despite a higher infant mortality in the cystic fibrosis group.

The possibility that the frequency of the gene for cystic fibrosis is increasing through selection is discussed.

REFERENCES

DANKS, D. M., ALLAN, J., AND ANDERSON, C. M. 1965. A genetic study of fibrocystic disease of the pancreas. *Ann. Hum. Genet.* (Lond.) 28: 323–356.

HALLETT, W. Y., KNUDSON, A. G., AND MASSEY, F. J. 1965. Absence of detrimental effect of the carrier state of the cystic fibrosis gene. *Ann. Rev. Resp. Dis.* 92: 714–724.