Am. J. Hum. Genet. 42:808–815, 1988

A Test of the Heterozygote-Advantage Hypothesis in Cystic Fibrosis Carriers

L. B. Jorde* and G. M. Lathrop†

*Department of Human Genetics and †Howard Hughes Medical Institute, University of Utah School of Medicine, Salt Lake City

Summary

We report a test of the hypothesis that the high frequency of cystic fibrosis (CF) in Caucasian populations is due to a fertility advantage in CF carriers. One hundred forty-three grandparent couples of Utah CF cases were compared with 20 replicate sets of matched control couples drawn from the Utah Genealogical Database. Ascertainment correction, which has not been applied in previous studies of CF carrier fertility, was applied to these data. Before ascertainment correction was applied, CF carriers appeared to manifest a significant fertility advantage over controls. After the correction formula was applied, this difference disappeared. Carriers and controls were also compared in terms of the length of intervals between births. Again, no significant differences were found. It was concluded that fertility differences are unlikely to account for the observed Caucasian CF gene frequency. Other mechanisms, particularly a past selective event or random genetic drift, are more likely to be responsible.

Introduction

Cystic fibrosis (CF) is routinely cited as the most common lethal genetic disease in Caucasian populations, with most estimates of prevalence at birth ranging from 1/2,000 to 1/4,000 (Harris and Nadler 1983). A variety of mechanisms, including genetic drift (Wright and Morton 1968), multiple loci (Schaap and Cohen 1976), high mutation rate (Goodman and Reed 1952), replacement after the birth of a CF child (Burdick 1977; Edwards 1977), and heterozygote advantage (Danks et al. 1965), have been proposed to account for this high frequency.

Of these explanations, heterozygote advantage has received the most attention from investigators. Several studies have reported differences in fertility between CF carriers and controls (Danks et al. 1965; Knudson et al. 1967; Conneally et al. 1973). The fertility of grandparents of CF probands was evaluated in these analyses, since each couple nearly always includes one gene carrier. The couple's reproductive behavior should not be altered because they typically would not produce affected offspring. Two major criticisms have been directed at these studies. First, the selection of controls has not been optimal in some cases (Mayo 1970). Second, none of the analyses has corrected adequately for ascertainment bias: those carrier grandparents who are more fertile are more likely to produce affected grandchildren (Wright and Morton 1968; Mayo 1970; Ten Kate 1977).

These difficulties were overcome in the present study. The study population consisted of grandparents of probands taken from a Utah registry of CF patients, and the control population consisted of well-matched replicate samples drawn from the computerized Utah Genealogical Database (Skolnick 1980). A correction for ascertainment bias has been formulated and applied to the data.

In addition to assessing case-control differences in family size, differences in birth intervals were analyzed. Although previous studies of heterozygote advantage in CF have not addressed this issue, it is possible that an elevated gene frequency could be due to

Received November 16, 1987.

Address for correspondence and reprints: Dr. L. B. Jorde, Department of Human Genetics, 501 Wintrobe Building, University of Utah School of Medicine, 50 North Medical Drive, Salt Lake City, UT 84132.

0002-9297/88/4206-0002$02.00

shorter birth intervals among the offspring of carriers even in the absence of a family size difference. Reduced birth intervals, which could reflect increased fecundity in CF carriers, would result in shorter average generation lengths in carrier families. Over time, this would increase the frequency of the CF gene.

Material and Methods

A total of 228 CF probands were ascertained on the basis of the clinical records of the Intermountain Cystic Fibrosis Center, Salt Lake City. The families of 44 probands were inaccessible owing to changes of address. The remaining 184 families were each sent both a letter explaining the study and a questionnaire requesting a 3-generation descending genealogy beginning with the proband's grandparents (generation 1). Information was thus obtained for all first-, second-, and third-degree relatives of each subject in the subject's own generation and in his or her two ascending generations. Families were asked to provide names of each individual as well as dates and places of birth, marriage, and death. Also, adoption and disease status (affected or not affected with CF) were given for each member of generations 2 and 3. Religious preference was indicated for each member of generations 1 and 2.

Six weeks after the first mailing, the families that did not respond were contacted again by mail. Nonrespondents to both mailings were then contacted by telephone. Thirteen of 184 families refused to participate in the study. An additional 14 families submitted information that was either incomplete or unusable for other reasons (e.g., the proband had been adopted). Forty families indicated willingness to participate but never submitted completed genealogies. Thus, the response rate, including the 14 families who submitted incomplete information, was 131/184 (71%). Excluding these 14 families, the response rate was 117/184 (64%).

The 117 responding families provided 234 grandparent couples ("carrier couples") for analysis. Since the control population was drawn from the Utah Genealogical Database, which consists almost exclusively of Mormons (Church of Jesus Christ of Latter-Day Saints), analysis was restricted to those carrier couples who had at least one member claiming affiliation with the Mormon Church. To avoid truncation of family size due to death of one of the grandparents, only those couples in which both grandparents survived beyond age 50 years were included. The analysis presented below is based on the 143 carrier couples ("cases") who met these criteria.

The Utah Genealogical Database consists of >1.2 million individuals linked by computer into large genealogies. Names, as well as dates and locations of birth, death, and marriage, are known for most members of the data base (further details are given in Skolnick [1980]). In the data base all couples were identified who matched a given carrier couple on the following characteristics: grandfather's birth decade, grandmother's birth year (within 5 years), and date of marriage (within 5 years). As in the carrier couples, both members of the control couple had to have survived beyond age 50 years and at least one member of the control couple had to have been baptized in the Mormon Church. Childless couples and adopted offspring of carrier and control couples were excluded. The matching process produced populations of potential control couples (generally on the order of several hundred to several thousand) for each carrier couple. To avoid spurious results in case-control comparisons, 20 replicate control couples were drawn randomly from within these populations for each carrier couple.

The average number of offspring born to the carrier couples was compared with the average number born to each of the 20 replicate control groups. Differences in these numbers were assessed statistically using *t*-tests. Since family size in humans tends to follow a negative binomial distribution (Brass 1958), the *t*-test is not entirely appropriate in this application. Thus, a nonparametric median test (Siegel 1956) was also used to test these differences. These tests were also used to evaluate average case-control differences in the intervals between marriage and first birth and between subsequent births.

The effect of ascertainment bias on case-control fertility differences was estimated using the following:

$$Q(n) = \frac{P(n)\left[\sum_{m=1}^{n} [1 - (1 - q)^m] \binom{n}{m} (0.5)^n\right]}{S}, \quad (1)$$

where $P(n)$ is the empirically observed distribution of family sizes in controls (the 20 replicate sets were pooled) and q is the probability that a CF heterozygote produces an affected child (assuming a carrier frequency of 1/25, $q = .01$). The term $\binom{n}{m}(0.5)^n$ specifies the probability that, of the grandparents' n offspring, m are heterozygotes. The term [1 –

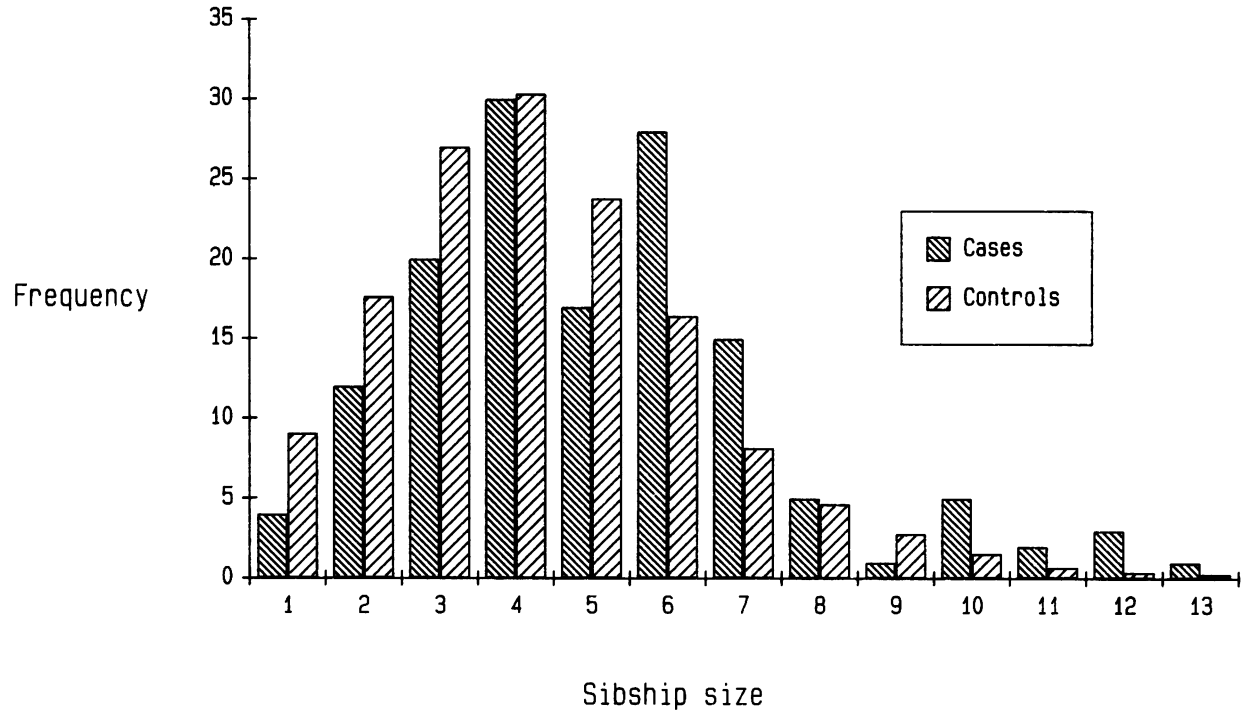

Figure 1 Distribution of family sizes in carrier and control couples. The control distribution represents averages over 20 replicate sets of controls.

$(1-q)^m$] gives the probability that, among these m heterozygotes, one or more will produce an affected child. These probabilities are then summed over all values of m. To convert $Q(n)$ into a probability density function, the numerator of equation (1) is divided by S, which is simply the numerator summed over n. $Q(n)$, then, is the adjusted control family-size distribution that is compared with the family-size distribution observed among the cases.

This adjustment is similar in concept to the one suggested by Ten Kate (1977). However, Ten Kate's correction was formulated for a study design in which control families are selected and matched on the basis of characteristics of the affected probands, rather than on the basis of characteristics of grandparents. Thus, the correction factor given above is more appropriate for this study design. To illustrate the effect of ascertainment correction, family-size comparisons are reported first without and then with the correction.

Results

Figure 1 shows the sibship-size distributions for the offspring of carrier and control couples; the distributions for the latter represent the average over the 20 replicate control sets. The average number of offspring born to carrier couples was 5.15, and the overall average born to the control couples, without adjustment for ascertainment bias, was 4.32. In each of the 20 comparisons of carrier couples with replicate controls, the average for carriers exceeded that for the controls. When t-tests were used, this difference was significant at the .01 level in 18 of 20 comparisons. When the nonparametric median test was used, the difference was significant at the .05 level in 12 of 20 comparisons.

The distributions of intervals between marriage and first birth are shown in figure 2. Again, the frequencies for control couples represent averages over 20 replicates. In both the carriers and controls, ~15% of the couples had intervals of <40 wk. These couples were omitted from this analysis. The average interval for carrier couples was 634 days, while that of the control couples was 709 days. The control interval exceeded that of the carriers in 19 of 20 comparisons. If the replicate control values are considered to be a rough empirical distribution, this would correspond to a .05 significance level. However, the difference in intervals was statistically significant

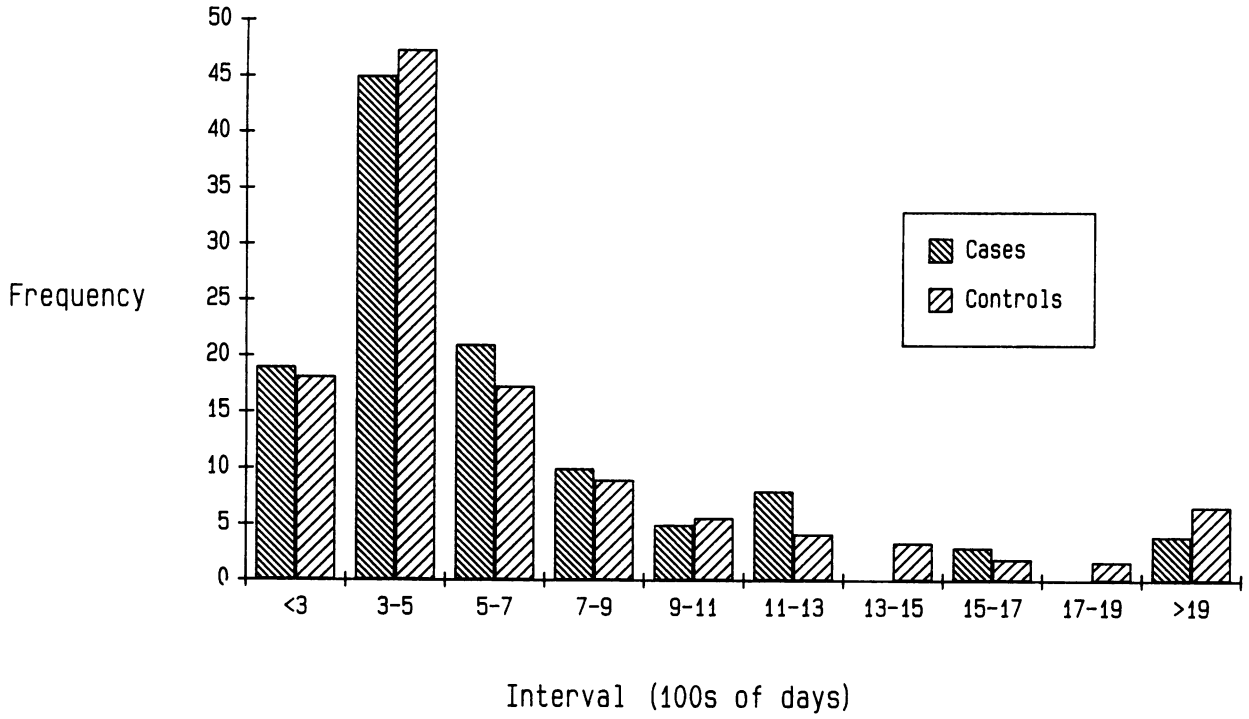

Figure 2 Distribution of the intervals, in hundreds of days, between marriage and the birth of the first offspring, in carrier and control couples (the latter were averaged over 20 replicate sets).

($P < .02$) by t-test in only one comparison, and it was not significant in any of the nonparametric tests. Thus, differences in this interval, although discernible, are marginally significant at best.

Figure 3 gives the distribution of intervals between births by parity level (level 2 indicates the interval between first and second births, and so on). Overall, the average birth interval for carrier couples was 1,103 days, while that of control couples was 1,155 days. The control couples' intervals exceeded those of carriers in 19 of 20 comparisons. This difference was significant ($P < .05$) in four of 20 comparisons when t-tests were used, but it was significant ($P < .01$) in only one comparison when the median test was used. Figure 3 indicates that case-control differences did not follow any regular pattern with respect to parity level.

Family-size differences in carrier couples and controls were evaluated after applying equation (1) to the control distribution, $P(n)$. The resulting distribution, $Q(n)$, is shown in figure 4, along with the previously shown distributions of carriers and controls. This adjusted control group distribution yielded an average family size of 5.33 for controls, which is slightly *greater* than the average carrier family size of 5.15. Thus, correction for ascertainment bias erased an apparently significant fertility difference between carriers and controls. It is likely that the observed birth-interval differences, which were not impressive, were due to this same ascertainment bias (i.e., larger families would tend to have shorter birth intervals).

Discussion

The results of the present study argue against higher fertility among CF carriers as a cause of the high frequency of the CF gene in Caucasians. The often-cited study by Knudson et al. (1967) did find a significant fertility difference between carrier couples (i.e., grandparents of CF cases) and controls (4.34 vs. 3.43 children). The control population in that study was not well defined or matched. Interestingly, the magnitude of this fertility difference is similar to that found in the present study *before* application of an ascertainment-correction factor. Danks et al. (1965) compared family sizes for CF grandparents with those for three sets of matched controls. Significant differences were found with only one of the control sets (4.6 in carrier couples vs. 4.0 in controls). Conneally et al. (1973) found conflicting evidence, with

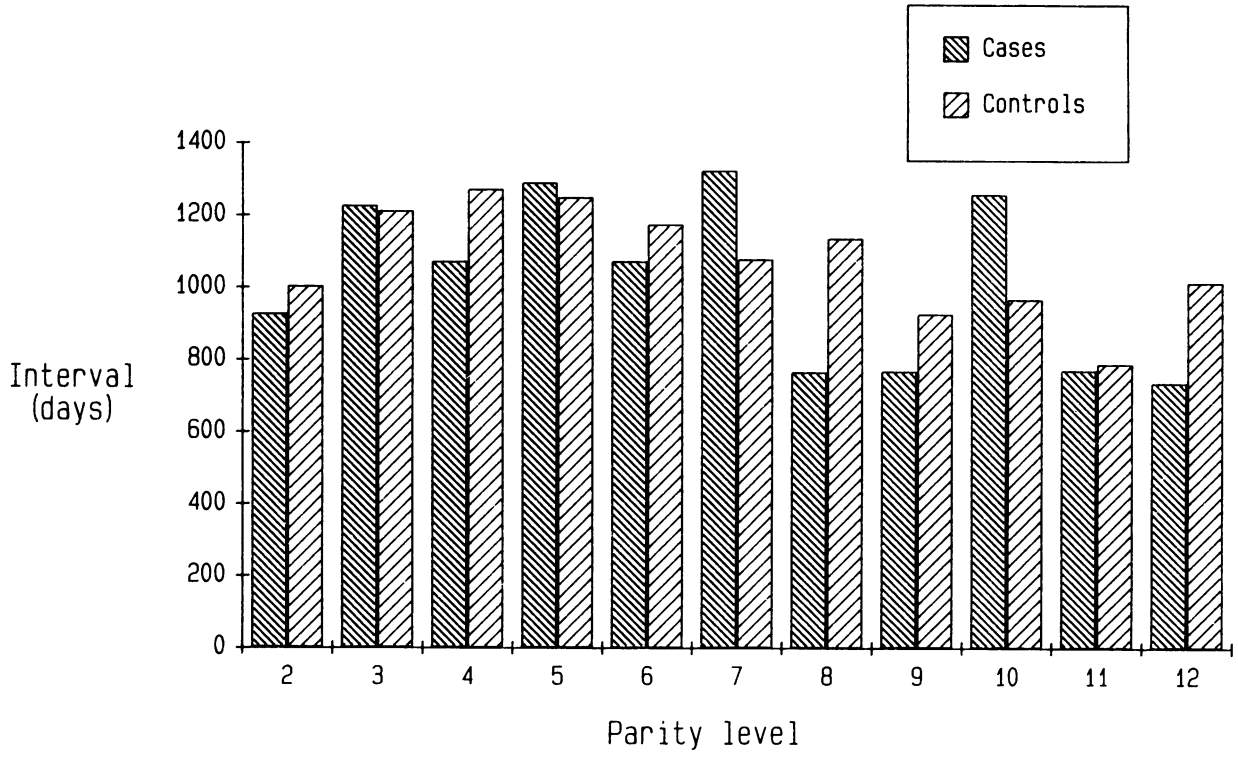

Figure 3 Distribution of average birth intervals, by parity, in carrier and control couples (the latter were averaged over 20 replicate sets).

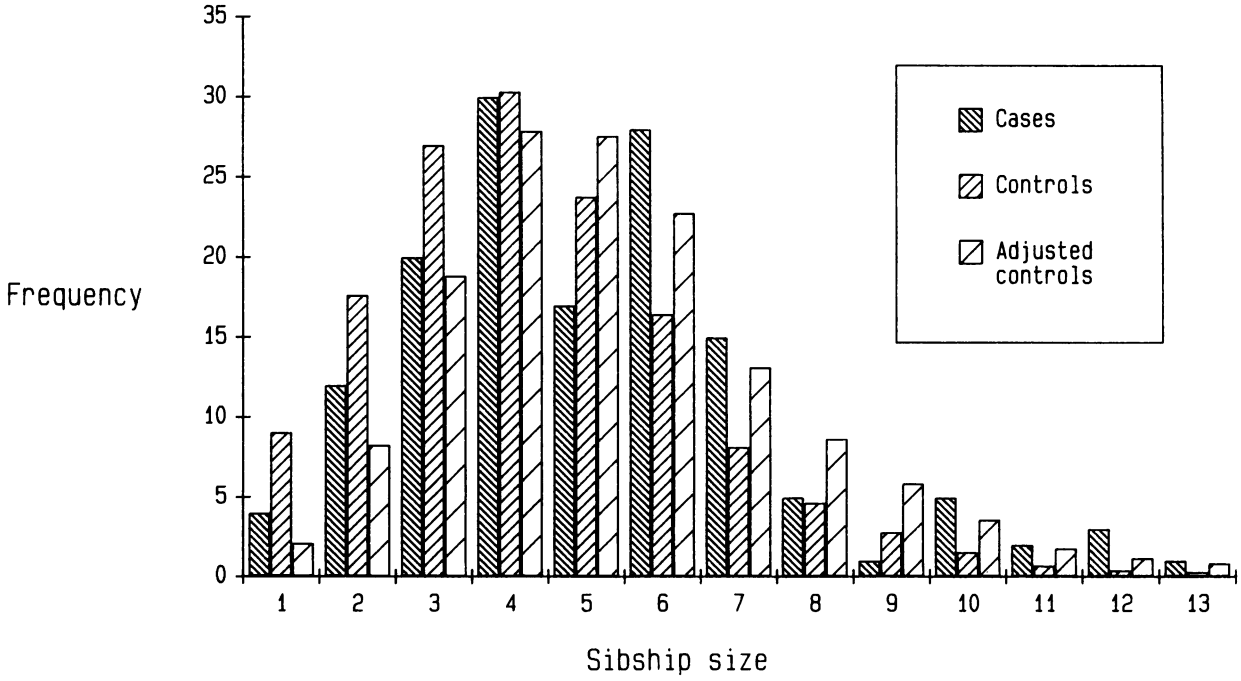

Figure 4 Distribution of family sizes in carrier and control couples, *including* the expected distribution of control family sizes after applying eq. (1).

carrier couples having larger family sizes than did controls in some decades but smaller family sizes in others. The results of these studies, already somewhat ambiguous, would be even less supportive of a fertility difference if an ascertainment-correction factor had been applied. Thus, there appears to be very little evidence in favor of a heterozygote advantage that is due to higher family sizes in CF carriers.

A weakness of the approach employed in the present study is that it is not known which grandparent carried the CF gene. If a fertility advantage were confined to only one sex (Anderson et al. 1966; Pritchard et al. 1983), half of the couples in the sample would not be expected to show an advantage. Direct carrier detection by using DNA polymorphisms now makes it possible to circumvent this problem. However, a major concern in using DNA polymorphisms for this purpose involves the sample size required to detect a reasonable fertility difference. It has often been estimated that, *at equilibrium,* a selective advantage of only ~2% could maintain CF at its present frequency (Steinberg and Brown 1960; Conneally et al. 1973). If one assumes that the mean and variance of family size are four (as in the Utah data), a sample size of more than 21,000 couples in each group would be required to detect a 2% fertility difference with a type I error level of .05 and a type II error level of .10 (Sokal and Rohlf 1981). This calculation assumes that a *t*-distribution can be appropriately used to test family-size differences; even larger sample sizes could be required if other distributions were used. Since the present study indicates that little, if any, difference is to be expected between carriers and controls in terms of fertility, it would appear inadvisable to attempt to gain additional information by using direct carrier diagnosis.

Heterozygote advantage has also been measured in terms of segregation distortion in CF families. Using the MET and pJ3.11 polymorphisms, Bowcock et al. (1986) determined carrier status in 228 unaffected siblings of CF cases. They found no significant deviation from the expected approximately 2:1 carriers:normals ratio. Similarly, Schmidtke et al. (1987) used four RFLPs to determine carrier status in 105 unaffected siblings of CF patients. Although the percentage of carriers (71%) was slightly greater than the expectation of 2/3, the difference was not statistically significant.

While the present study was concerned with a direct determination of fertility differences in carriers and controls, heterozygote advantage could also be mediated through a lowered mortality rate in CF carriers. It has been suggested that CF carriers might benefit from increased resistance to tuberculosis (Crawfurd 1972; Meindl 1987), typhus (Stuart and Burdon 1974), influenza (Shier et al. 1979), malaria (Super and Schalkwyk 1979), bubonic plague (Cassano 1985), or venereal syphilis (Hollander 1982). Although some interesting physiological mechanisms are proposed, the evidence in support of these hypotheses is only circumstantial. Actual tests (see, e.g., Hallett et al. 1965; Super and Schalkwyk 1979) have produced negative results.

The hypothesis that multiple CF loci may be responsible for the disease's high frequency has not been supported by linkage analyses, all of which indicate a lack of genetic heterogeneity for CF (Beaudet et al. 1986; Klinger et al. 1986; Vitale et al. 1986; Watkins et al. 1986; White 1986). In fact, ≥98% of the analyzed CF cases appear to be due to lesions at the same locus (Bowcock et al. 1986). Tests comparing CF prevalence among first cousins of probands with that among the general population have produced somewhat more equivocal results (Crow 1965; Conneally et al. 1973; Danks et al. 1983, 1984; Gedschold et al. 1987), as have studies based on consanguinity (Romeo et al. 1985, 1986; Lander and Botstein 1986). In part, this reflects the relative lack of resolving power of these methods. In any case, there is little evidence that multiple loci could be responsible for the high frequency of CF in Caucasians.

An elevated mutation rate at the CF locus also appears to be an implausible explanation, since it would require an extraordinarily high mutation rate in order to maintain CF at its current frequency. In addition, it seems unlikely that the mutation rate would be elevated only in some Caucasians and not in blacks or Asians (Steinberg and Brown 1960).

Wright and Morton (1968) estimated that there is a probability of ~.001 that the CF gene frequency could have reached its current value in Caucasians through genetic drift. While this probability is low, it would predict that roughly one of the hundreds of lethal recessive diseases identified in humans would in fact reach such a high frequency. CF could be the disease that, by chance, did so. In connection with the drift hypothesis, it is important to note that only *some* Caucasian populations have a high frequency of CF.

It is equally difficult to exclude the possibility of a transient selective event sometime in the recent history of this population. Given the slow rate at which

natural selection operates against a recessive gene—even when the latter is lethal in all cases—a past event involving selection or genetic drift could quite easily account for the present CF gene frequency (see Knudson et al. [1967] and Meindl [1987] for sample calculations). If a past selective event is the causal mechanism, elucidation of the biochemical defect responsible for CF, could shed light on the nature of this event.

Acknowledgments

We are grateful to Ann Wechsler and Carol Wood for technical assistance. Drs. Richard Meindl, Kenneth Morgan, Elizabeth O'Brien, Dorian Pritchard, and Ray White provided aid and advice. We also wish to express appreciation to the CF families who provided data for this study. This research was supported by a grant from the Cystic Fibrosis Foundation and by the Howard Hughes Medical Institute.

References

Anderson, C. M., J. Allan, and P. G. Johansen. 1966. Comments on the possible existence and nature of a heterozygote advantage in cystic fibrosis. Mod. Probl. Pediatr. **10**:381–387.

Beaudet, A., A. Bowcock, M. Buchwald, L. Cavalli-Sforza, M. Farrall, M.-C. King, K. Klinger, J.-M. Lalouel, G. Lathrop, S. Naylor, J. Ott, L.-C. Tsui, B. Wainwright, P. Watkins, R. White, and R. Williamson. 1986. Linkage of cystic fibrosis to two tightly linked DNA markers: joint report from a collaborative study. Am. J. Hum. Genet. **39**:681–693.

Bowcock, A. M., J. Crandall, L. Daneshvar, G. M. Lee, B. Young, V. Zunzunegui, C. Craik, L. L. Cavalli-Sforza, and M.-C. King. 1986. Genetic analysis of cystic fibrosis: linkage of DNA and classical markers in multiplex families. Am. J. Hum. Genet. **39**:699–706.

Brass, W. 1958. Models of birth distributions in human populations. Bull. Inst. Int. Stat. **36**:165–178.

Burdick, A. B. 1977. Frequency of the gene for cystic fibrosis with a view of replacement and recognition effects and reproduction by homozygotes. Hum. Hered. **27**:366–371.

Cassano, W. F. 1985. Cystic fibrosis and the plague. Med. Hypotheses **18**:51–52.

Conneally, P. M., A. D. Merritt, and P.-L. Yu. 1973. Cystic fibrosis: population genetics. Tex. Rep. Biol. Med. **31**:639–650.

Crawfurd, M. D. 1972. A genetic study, including evidence for heterosis, of cystic fibrosis of the pancreas. Heredity **29**:126.

Crow, J. F. 1965. Problems of ascertainment in the analysis of family data. Pp. 23–44 *in* J. V. Neel, M. W. Shaw, and W. J. Schull, eds. Genetics and epidemiology of chronic diseases. U.S. Government Printing Office, Washington, DC.

Danks, D. M., J. Allan, and C. M. Anderson. 1965. A genetic study of fibrocystic disease of the pancreas. Ann. Hum. Genet. **28**:323–356.

Danks, D. M., J. Allan, P. D. Phelan, and C. Chapman. 1983. Mutations at more than one locus may be involved in cystic fibrosis—evidence based on first-cousin data and direct counting of cases. Am. J. Hum. Genet. **35**:838–844.

Danks, D. M., P. D. Pheland, and C. Chapman. 1984. Retraction: *no* evidence for more than one locus in cystic fibrosis. Am. J. Hum. Genet. **36**:1401–1402.

Edwards, J. H. 1977. Heterozygote advantage. Arch. Dis. Child. **52**:343–344.

Gedschold, J., S. Kropf, R. Szibor, and M. Berger. 1987. Cystic fibrosis—a single locus disease? results of a population genetics study. Hum. Genet. **75**:277–280.

Goodman, H. O. and S. C. Reed. 1952. Heredity of fibrosis of the pancreas: possible mutation rate of the gene. Am. J. Hum. Genet. **4**:59–71.

Hallett, W. Y., A. G. Knudson, and F. J. Massey. 1965. Absence of detrimental effect of the carrier state for the cystic fibrosis gene. Am. Rev. Respir. Dis. **92**:714–724.

Harris, C. J., and H. L. Nadler. 1983. Incidence, genetics, heterozygote, and antenatal detection of cystic fibrosis. Pp. 1–7 *in* J. D. Lloyd-Still, ed. Textbook of cystic fibrosis. John Wright, Boston.

Hollander, D. H. 1982. Etiogenesis of the European cystic fibrosis polymorphism: heterozygote advantage against venereal syphilis? Med. Hypotheses **8**:191–197.

Klinger, K., P. Stanislovitis, N. Hoffman, P. C. Watkins, R. Schwartz, R. Doherty, P. Scambler, M. Farrall, R. Williamson, and B. Wainwright. 1986. Genetic homogeneity of cystic fibrosis. Nucleic Acids Res. **14**:8681–8686.

Knudson, A. G., L. Wayne, and W. Y. Hallett. 1967. On the selective advantage of cystic fibrosis heterozygotes. Am. J. Hum. Genet. **19**:388–392.

Lander, E. S., and D. Botstein. 1986. Consanguinity and heterogeneity: cystic fibrosis need not be homogeneous in Italy. Am. J. Hum. Genet. **39**:282–283.

Mayo, O. 1970. On the maintenance of polymorphisms having an inviable homozygote. Ann. Hum. Genet. **33**:307–317.

Meindl, R. S. 1987. Hypothesis: a selective advantage for cystic fibrosis heterozygotes. Am. J. Phys. Anthropol. **74**:39–45.

Pritchard, D. J., G. R. Hickman, and R. Nelson. 1983. Sex ratio and heterozygote advantage in cystic fibrosis families. Arch. Dis. Child. **58**:290–293.

Romeo, G., M. Bianco, M. Devoto, P. Menozzi, G. Mastella, A. M. Giunta, C. Micalizzi, M. Antonelli, A. Battistini, F. Santamaria, D. Castello, A. Marianelli, A. G.

Marchi, A. Manca, and A. Miano. 1985. Incidence in Italy, genetic heterogeneity, and segregation analysis of cystic fibrosis. Am. J. Hum. Genet. **37**:338–349.

Romeo, G., M. Devoto, and M. Bianco. 1986. Homogeneity vs. heterogeneity of cystic fibrosis in Italy. Am. J. Hum. Genet. **39**:283–284.

Schaap, T., and M. M. Cohen. 1976. A proposed model for the inheritance of cystic fibrosis. Pp. 291–303 *in* J. A. Mangos and R. C. Talamo, eds. Cystic fibrosis: projections into the future. Stratton Intercontinental, New York.

Schmidtke, J., and 29 others. 1987. Linkage relationships and allelic associations of the cystic fibrosis locus and four market loci. Hum. Genet. **76**:337–343.

Shier, W. T. 1979. Increased resistance to influenza as a possible source of heterozygote advantage in cystic fibrosis. Med. Hypotheses **5**:661–668.

Siegel, S. 1956. Nonparametric statistics. McGraw-Hill, New York.

Skolnick, M. 1980. The Utah Genealogical Database: a resource for genetic epidemiology. Pp. 285–297 *in* J. Cairns, J. L. Lyon, and M. H. Skolnick, eds. Banbury Report 4: cancer incidence in defined populations. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry: the principles and practice of statistics in biological research. W. H. Freeman, San Francisco.

Steinberg, A. G., and D. C. Brown. 1960. On the incidence of cystic fibrosis of the pancreas. Am. J. Hum. Genet. **12**:416–424.

Stuart, A. B., and M. G. Burdon. 1974. Frequency of the cystic-fibrosis gene. Lancet 2:1521.

Super, M., and D. J. van Schalkwyk. 1979. Heterozygote advantage in cystic fibrosis: mosquito tests. Clin. Genet. **16**:65–68.

Ten Kate, L. P. 1977. A method for analysing fertility of heterozygotes for autosomal recessive disorders, with special reference to cystic fibrosis, Tay-Sachs disease and phenylketonuria. Ann. Hum. Genet. **40**:287–297.

Vitale, E., M. Devoto, G. Mastella, and G. Romeo. 1986. Homogeneity of cystic fibrosis in Italy. Am. J. Hum. Genet. **39**:832–836.

Watkins, P. C., R. Schwartz, N. Hoffman, P. Stanislovitis, R. Doherty, and K. Klinger. 1986. A linkage study of cystic fibrosis in extended multigenerational pedigrees. Am. J. Hum. Genet. **39**:735–743.

White, R. L. 1986. The search for the cystic fibrosis gene. Science **234**:1054–1055.

Wright, S. W., and N. E. Morton. 1968. Genetic studies on cystic fibrosis in Hawaii. Am. J. Hum. Genet. **20**:157–169.