

Segregation Analysis of Juvenile Diabetes Mellitus*

ITALO BARRAI and HOWARD M. CANN†

From the International Laboratory of Genetics and Biophysics, Pavia Section, and the Institute of Genetics, University of Pavia, Italy

Twin studies (Joslin, Root, White, and Marble, 1952) and analyses of pooled family data (Pincus and White, 1934; Simpson, 1962) provide evidence for a genetic aetiology of diabetes mellitus. The wide variation in the age at onset and in severity of the disease and the possibility that the disease is genetically heterogeneous present serious obstacles to establishing a mode of inheritance. Nevertheless, results of a number of studies have suggested that the disease is determined by a recessive gene at a single autosomal locus (Pincus and White, 1934; Steinberg and Wilder, 1952; Post, 1962).

The techniques of segregation analysis can correct for biased ascertainment and sporadic cases (phenocopies, mutations, and other isolated events of non-segregational origin) in samples of families (Bailey, 1951; Morton, 1958, 1959). Under multiple selection, the segregation frequency and other pertinent genetic parameters may be estimated using maximum likelihood scores, provided that probands and ascertainment are counted for each sibship. From 233 families containing one or more juvenile diabetics, Simpson (1962) has published an unusually good collection of data that meets the requirements for segregation analysis. We have applied maximum likelihood segregation analysis to Simpson's data to test the hypothesis that diabetes mellitus is caused by a recessive gene at an autosomal locus. Our analysis is presented here.

The Data

For each family ascertained Simpson has recorded the following:

- (1) Number of probands.
- (2) Number of affected sibs.

- (3) Number of normal sibs.
- (4) Age at onset of diabetes mellitus.
- (5) Age at death or last examination of affected and unaffected sibs.
- (6) Diabetes status of parents, their sibs, and of grandparents.

There had been 223 probands in the Hospital for Sick Children, Toronto, and 17 had been referred to Simpson by another Toronto hospital, by private physicians, and through the Canadian Diabetic Association; 7 families were ascertained more than once. The probands were selected for their clinical homogeneity (i.e. juvenile diabetes), most having developed diabetes at the age of 14 years or less. The average age of parents of probands was about 60 years, and only 8% of parents were younger than 40.

Assumptions

- (1) Predisposition to diabetes mellitus is inherited.
- (2) The distribution of glucose tolerance tests or some other parameter accurately representing the genotype is bimodal or, perhaps, trimodal. To our knowledge, such distributions have not been published.
- (3) The probability of any diabetic individual being a proband is constant and independent of that for any other diabetic in the families reported by Simpson. For any model of ascertainment this assumption does not accurately represent actual conditions, but if the variation in the probability of ascertainment for a given sample is small the bias will not significantly affect the estimates of genetic parameters.

The Hypothesis to be Tested

Recessive transmission at a single autosomal locus is the null hypothesis. If the predisposition for diabetes mellitus is determined by a fully penetrant allele in homozygous state and there are no sporadic cases, the segregation frequency in intercrosses will be 0.25; in backcrosses the segregation frequency will be twice this value.

As an alternative hypothesis, we may assume incomplete penetrance, which will affect the segregation frequency; in this case as well, the segregation frequency in backcrosses will be twice that in intercrosses.

If there are sporadic cases, the segregation frequency among offspring of unaffected parents will be less than one-half the value obtained from backcrosses.

Received October 26, 1964.

* Supported by Euratom-C.N.R.-C.N.E.N. Contract no. 012-61-12 BIAI.

† Senior Postdoctoral Fellow, N.S.F., 1962-1963. Present address: Department of Pediatrics, Stanford University, School of Medicine, Palo Alto, California, U.S.A.

TABLE I

SUMMARY OF SIMPSON'S (1962) DATA: DISTRIBUTION OF DIABETIC SIBSHIPS BY SIBSHIP SIZE AND BY NUMBER OF AFFECTED AND NUMBER OF PROBANDS OF UNAFFECTED PARENTS (INTERCROSSES)

	Sibship Size												
	1	2	3	4	5	6	7	8	9	11	12	13	
	Number of Sibships												
One affected													
One proband	26	52	47	31	11	14	4	3	1	1	1	1	
Two affected													
One proband		3	3	3	3	2	1						
Two probands		1	3	1	1								
Three affected													
One proband			1							1			
Two probands					1								
Total	26	56	54	35	15	17	5	3	1	1	1	1	

From Simpson's data, we include only families in which the proband is an offspring, and there are 204 such families giving relevant information—190 matings with normal parents, which we assume to be intercroses (Table I), and 14 with one parent affected, which we assume to be backcrosses (Table II). We are cautious in classifying matings as intercroses because of the possibility of delayed age at onset among parents. However, age distributions for parents of all probands in the sample indicate that many of the unaffected parents who have attained a high risk age will not develop diabetes mellitus, and we presume that they are heterozygotes.

Methods of Analysis

The theory, probability formulae, and maximum likelihood scores and variances for the segregation analysis have been presented by Morton (1958, 1959). A maximum likelihood score, u_0 , which evaluates the fit of the parameter θ , is computed for each sibship:

$$u_0 = \frac{d \ln P}{d\theta}$$

where P represents the pertinent probability distribution for the sibship. When the parameter(s) being estimated give a perfect fit, $u_0 = 0$. Scores are additive so that $\Sigma u_0 = U_0$. The variance of the maximum likelihood score is:

$$k_0 = -E \frac{d^2 \ln P}{d\theta^2}, \text{ and } \Sigma k_0 = K_{00}$$

U_0^2 / K_{00} is distributed as χ^2 with one degree of freedom and tests the fit for θ . The equations for maximum likelihood scores and variances were programmed in FORTRAN* (McCracken, 1961) for an Olivetti 'Elea' digital computer which was used for the analysis.

* FORTRAN Programming System, International Business Machines Corporation, Data Processing Division, White Plains, New York, U.S.A.

TABLE II

SUMMARY OF SIMPSON'S (1962) DATA: DISTRIBUTION OF DIABETIC SIBSHIPS BY SIBSHIP SIZE AND BY NUMBER OF AFFECTED AND NUMBER OF PROBANDS ONE PARENT AFFECTED (BACKCROSSES)

	Sibship Size							
	1	2	3	4	5	6	7	8
	Number of Sibships							
One affected								
One proband	2	4	1	2	2	1		
Two affected								
One proband			1			1	1	1
Total	2	4	1	3	2	2	1	1

From the distribution of probands among affected members of sibships, it is possible to estimate the probability of ascertainment, π , that any affected individual be a proband (Morton, 1959). In sibships with r affected members, the distribution of a probands ($a > 0$) is:

$$P(a) = \binom{r}{a} \pi^a (1 - \pi)^{r-a} / [1 - \pi]^r \tag{Problem 9}$$

Morton and Chung (1959) refer to the solution as this distribution by maximum likelihood scores of problem 9, and we will use their terminology. From problem 9, π for intercross sibships was estimated as 0/394, and this was the value used in problems 2 and 3 to estimate segregation frequency for intercroses. To estimate segregation frequency for backcrosses with one or more affected offspring, the same value of π was used because none of the parents was a proband. If such had been the case, π should have been fixed at 1, ascertainment then being equivalent to truncate selection (Morton, 1959).

The frequency distribution for sibships of various sizes each containing one affected member ($r = 1$), i.e. for simplex sibships, is:

$$P(r = 1 | r > 0) = \frac{sp \pi [x + (1 - x) q^{s-1}]}{xsp \pi + (1 - x) [1 - (1 - p \pi)^s]} \tag{Problem 2}$$

where p is the segregation frequency, $q = 1 - p$, s is the sibship size, and x is the relative frequency of sporadic cases among affected cases. Solution by maximum likelihood scores permits an estimation of segregation frequency. This genetic parameter is also estimated from the maximum likelihood solution of the distribution of probabilities for sibships with two or more affected members ($r > 1$), i.e. multiplex families:

$$P(r | r > 1) = \frac{\binom{s}{r} Prq^{s-r} [1 - (1 - \pi)^r]}{1 - (1 - p\pi)^s - \pi^s p q^{s-1}} \tag{Problem 3}$$

Results and Discussion

Table III presents a summary of the segregation analysis of Simpson's data assuming complete penetrance and absence of sporadic cases ($x = 0$). χ^2 evaluation of each of the parameters, x , π , and p , for intercross and backcross data indicates that one or both assumptions are untenable, i.e. there is segregation distortion due to reduced penetrance or an excess of sporadic cases.

A summary of the estimation of segregation frequency in both mating types is presented in Table IV; again, absence of sporadic cases is

assumed. For intercrosses $p = 0.067$ and for backcrosses $p = 0.105$, the latter being almost twice the former value. The frequency of sporadic cases does not differ significantly from 0 in both mating groups ($\chi^2_{(1)} = 0.876$), indicating that segregation distortion is due to reduced penetrance. The segregation frequency is the product of the theoretical segregation frequency and the penetrance, y . Thus for intercrosses, $y = 0.067/0.250 = 0.268$, and for backcrosses, $y = 0.105/0.500 = 0.210$.

The heterogeneity between segregation frequency values for multiplex and simplex sibships arising

TABLE III
TEST FOR COMPLETE PENETRANCE AND ABSENCE OF SPORADIC CASES

Source	Problem	p	x	π	Up	Ux	U π	Kpp	Kpx	Kp π	Kxx	Kx π	K $\pi\pi$
<i>Presumptive Intercrosses</i>													
a > 0	2	0.25	0	0.394	-418.55	230.31	45.79	1712.79	-719.28	-189.83	412.08	78.21	21.34
a > 0	3	0.25	—	0.394	-16.35	—	1.33	190.35	—	-15.81	—	—	1.33
a > 0	9	—	—	0.394	—	—	0	—	—	—	—	—	93.53
Total					-434.90	230.31	47.12	1903.14	-719.28	-205.64	412.08	78.21	116.20
<i>Presumptive Backcrosses</i>													
a > 0	2	0.50	0	0.394	-42.96	72.44	6.02	64.46	-128.49	-9.09	525.00	16.86	1.29
a > 0	3	0.50	—	0.394	-24.32	—	2.56	71.94	—	-6.44	—	—	0.60
Total					-67.28	72.44	8.58	136.40	-128.49	-15.53	525.00	16.86	1.89

Tests of Fit of Estimates to the Data

Mating Type	Parameter Tested	Fixed Parameters	$\chi^2_{(1)}$
Intercrosses	$p = 0.250$ $x = 0$	$x = 0, \pi = 0.394$ $p = 0.250, \pi = 0.394$	99.382 128.719
Backcrosses	$\pi = 0.394$ $p = 0.5$ $x = 0$ $\pi = 0.394$	$x = 0, \pi = 0.394$ $p = 0.5, \pi = 0.394$ $p = 0.5, x = 0$	19.108 33.186 9.995 38.950

TABLE IV
TEST FOR REDUCED PENETRANCE AND ABSENCE OF SPORADIC CASES

Source	Problem	p	x	π	Up	Ux	U π	Kpp	Kpx	Kp π	Kxx	Kx π	K $\pi\pi$
<i>Presumptive Intercrosses</i>													
a > 0	2	0.067	0	0.394	-39.46	4.09	1.43	6311.80	-476.19	-240.12	36.39	18.09	9.14
a > 0	3	0.067	—	0.394	39.52	—	1.23	401.08	—	-12.08	—	—	0.37
a > 0	9	—	—	0.394	—	—	0	—	—	—	—	—	93.53
Total					0.06	4.09	2.66	6712.88	-476.19	-252.20	36.39	18.09	103.04
<i>Presumptive Backcrosses</i>													
a > 0	2	0.105	0	0.394	8.79	-1.61	-0.48	351.31	-46.46	-20.06	6.23	2.65	1.15
a > 0	3	0.105	—	0.394	-8.79	—	+0.40	93.61	—	-4.18	—	—	0.19
Total					0.00	-1.61	-0.08	444.92	-46.46	-24.24	6.23	2.65	1.34

Parameters tested $\chi^2_{(1)}$
 x between mating types 0.876
 p between problems, intercrosses 4.141
 p between problems, backcrosses 1.045

from intercrosses ($\chi^2_{(1)} = 4.141$) may indicate heterogeneity of penetrance. The segregation frequency estimated from problem 2 for simplex sibships was low and, assuming $x = 0$ (no sporadic cases), this low value must reflect low penetrance in these families.

Delayed age of onset is a component of penetrance especially important when considering diabetes mellitus. If $f(z)$ is the distribution of age at death or last examination for affected and unaffected sibs omitting the ages of probands first ascertained (index cases), and $G(z)$ is the cumulative distribution of age at onset of diabetes mellitus in the sample, then the average penetrance approximates:

$$y = \int f(z)G(z)dz = 0.489 \text{ (Morton, 1959),}$$

which, being the case for single selection, is limiting. Comparing this value of y to those obtained above from segregation analysis suggests that age-related factors contribute considerably, but not completely, to penetrance in this sample of juvenile diabetic families. There is no evidence of sporadic cases, and penetrance may be as low as 25%.

The finding of a segregation frequency for backcrosses which is almost twice that for intercrosses is consistent with the hypothesis being tested. This finding rests on analysis of small numbers of backcrosses and may prove anomalous with a larger sample. Further, we must point out some ambivalence of the data. The incidence of diabetes mellitus among sibs of probands is also consistent with that expected if the disease were determined by genes at multiple loci. Edwards (1960) has discussed the difficulties in differentiating between recessive inheritance with low penetrance and multifactorial inheritance. For a disorder which is the consequence of multifactorial inheritance, Edwards (1960) has shown that the incidence of the disorder in sibs of probands ($i_{c_{exp}}$) approximates the square root of the population incidence of the disorder, $\sqrt{i_p}$:

$$i_{c_{exp}} = \sqrt{i_p}$$

Population surveys for incidence of diabetes mellitus (Kenny, Chute, and Best, 1951; Kenny and Chute, 1953), cited by Simpson, permit estimates of i_p and $i_{c_{exp}}$:

$$i_p = 0.0056 \pm 0.0032$$

$$i_{c_{exp}} = 0.0747 \pm 0.0214.$$

The observed incidence of diabetes mellitus among sibs of probands ($i_{c_{obs}}$) is

$$i_{c_{obs}} = 0.0523 \pm 0.0096.$$

The observed and expected values for i_c do not differ significantly, indicating that the data fit the expectation for multifactorial inheritance.

In invoking low penetrance to explain the marked segregation distortions found in this analysis, we recognize that the phenomenon could completely, or in part, result from combined expressions of genes at multiple loci. An indicator accurately representing genotype should discriminate between single and multifactorial inheritance hypotheses for diabetes mellitus. Bimodal frequency distributions for such a measure will support single factor transmission.

Summary

Segregation analysis of 204 sibships containing one or more juvenile diabetics demonstrated: (1) a segregation frequency for backcross matings which is about twice this parameter in intercross matings; (2) penetrance as low as 25% in both mating groups; and (3) absence of sporadic cases. These findings support the hypothesis that diabetes mellitus is transmitted recessively at a single locus. However, the low penetrance may reflect combined expressions of genes at multiple loci. It should be noted that the data fit also the expectations for multifactorial transmission.

We thank Drs N. E. Morton, W. Bodmer, and L. L. Cavalli-Sforza for much helpful criticism during the preparation of this paper. The financial help of the U.S. Atomic Energy Commission is gratefully acknowledged.

REFERENCES

- Bailey, N. T. J. (1951). The estimation of the frequencies of recessives with incomplete multiple selection. *Ann. Eugen. (Lond.)*, **16**, 215.
- Edwards, J. H. (1960). The simulation of mendelism. *Acta genet. (Basel)*, **10**, 63.
- Joslin, E. P., Root, H. F., White, P., and Marble, A. (1952). *The Treatment of Diabetes Mellitus*, 9th ed., pp. 57-58. Lea and Febiger, Philadelphia.
- Kenny, A. J., and Chute, A. L. (1953). Diabetes in two Ontario communities; Studies in case finding. *Diabetes*, **2**, 187.
- , —, and Best, C. H. (1951). A study of the prevalence of diabetes in an Ontario community. *Canad. med. Ass. J.*, **65**, 233.
- McCracken, D. D. (1961). *A Guide to FORTRAN Programming*. John Wiley, New York.
- Morton, N. E. (1958). Segregation analysis in human genetics. *Science*, **127**, 79.
- (1959). Genetic tests under incomplete ascertainment. *Amer. J. hum. Genet.*, **11**, 1.
- , and Chung, C. S. (1959). Formal genetics of muscular dystrophy. *ibid.*, **11**, 360.
- Pincus, G., and White, P. (1934). On the inheritance of diabetes mellitus II. Further analysis of family histories. *Amer. J. med. Sci.*, **188**, 159.
- Post, R. H. (1962). An approach to the question, does all diabetes depend upon a single genetic locus? *Diabetes*, **11**, 56.
- Simpson, N. E. (1962). The genetics of diabetes: A study of 233 families of juvenile diabetics. *Ann. hum. Genet.*, **26**, 1.
- Steinberg, A. G., and Wilder, R. M. (1952). A study of the genetics of diabetes mellitus. *Amer. J. hum. Genet.*, **4**, 113.