Colchester revisited: A genetic study of mental defect*

NEWTON E. MORTON, D. C. RAO, HELEN LANG-BROWN, CHARLES J. MACLEAN,
ROBERT D. BART, and RUTH LEW

From Population Genetics Laboratory, University of Hawaii, Honolulu, Hawaii; The Kennedy-Galton Centre for Clinical Genetics, Harperbury Hospital, U.K.; and Pacific Institute of Rehabilitation Medicine, Honolulu, Hawaii

SUMMARY This reanalysis of a classic survey leads to inferences about design of genetic studies, resolution of heterogeneity, and the role of autosomal and sex-linked genes in mental retardation, which is no longer refractory to segregation analysis. By discriminating between sociofamilial and biological types we estimate that at least 351 autosomal loci can produce mental retardation, with an inbred load of 0.83 detrimental equivalents and a mutation rate of 0.008 per gamete, or less than 2.4×10^{-5} per locus. The distribution of probands was estimated as: 7 per cent medical, 60 per cent sociofamilial, and 33 per cent biological. Simple genetic mechanisms account for virtually all the biological category. Within the sociofamilial group cultural inheritance and polygenes could not be resolved.

In 1938 L. S. Penrose published 'A clinical and genetic study of 1280 cases of mental defect' from the Royal Eastern Counties Institution, Colchester, England. In the preface, the Medical Research Council predicted 'There can be no doubt that Dr Penrose's methods, data, and results will to a large extent determine the general course of research in mental deficiency in this country for some years to come.'

Although this anticipation proved conservative, genetic inferences were limited not only by the contemporary state of biological, clinical, and cytogenetic approaches, but also by restriction of genetic analysis to simple Mendelian entities. It was not until 1965 that Dewey *et al* were able to apply advances in segregation and consanguinity analysis to part of the Colchester study, 'yielding information about the number of loci and gene frequencies of recessive factors for severe mental defect, their penetrance and mutation rates, and the mechanism which maintains these genes in the population despite the greatly reduced fertility of affected persons'. Their analysis depended on segregation of family members into severely affected and nor-

mal, in frequencies that indicated a mixture of sporadic cases and recessive genes of high penetrance. They considered that 'overlap with normal makes the group of mild defectives unsuitable for segregation analysis'.

The last decade has seen extension of methods of family analysis under a model which includes incomplete ascertainment, a major locus, polygenes, and environment common to sibs, so that it is unnecessary to assume discrete segregation of normal and affected sibs (Morton, 1974). Mild mental defect is no longer refractory to genetic analysis, which we shall here apply to the Colchester exemplar.

Prevalence and incidence

Penrose recognized 6 intellectual grades for patients on the basis of Stanford-Binet and Porteus tests, while their relatives were rated by interview. 'The scale of assessment was intended to correspond with the grades into which the patients were divided' (Penrose, 1938, p. 18). For our analysis unclassified parents were taken as of unknown score, drawn randomly from the population, while unclassified sibs were omitted (Table I). Because grades are discrete we shall not regard them as a quantitative trait.

* This work was supported by Grants GM 17173 and HL 16774 from the U.S. National Institutes of Health.
Received for publication 12 January 1976

TABLE I
PREVALENCE IN THE COLCHESTER POPULATION

Grade	Code	I. Q range	No of Patients	Parents of Patients	Sibs of Patients	More Remote Relatives	Hospitalization Rate	Prevalence
Superior	S	115 +	—	11	68	82	—	—
Normal	N	85–114	—	1908	3645	10 138	—	—
Dull	M ₀	70–84	179	351	493	423	0.03	24 492
Simpleton	M ₁	50–69	448	191	272	181	}	3061
Imbecile	M ₂	20–49	433	4	117	50		
Idiot	M ₃	< 20	220	—	50	14		
Total			1280	2465 + 95 unknown	4645	10 888	—	27 553

Prevalence is defined as the frequency of cases living and certifiable in the general population. There were two million persons resident in the Eastern counties (Essex, Suffolk, and Cambridge-shire). Penrose (1963, p. 64) considered that only 25% of severe defectives and 3% of mild defectives were in hospital. Since the 1280 Colchester patients were a representative sample of 1500 patients (Penrose, 1938, p. 7), the estimated prevalence of mild defect is $(1500)(627)/(1280)(0.03) = 24\,492$ cases in a population of two million. For severe defect the prevalence is $(1500)(653)/(1280)(0.25) = 3061$ cases. Therefore in a population of 100 000 we expect 153 cases of severe mental defect and 1225 cases of mild defect.

Whereas prevalence considers only living certifiable cases, *incidence* is defined as the frequency among survivors to age 1 of persons who, at any time in their life, are certifiably retarded. Diagnosis of severe defect is virtually complete, but premature mortality reduces prevalence. Survival of mild defect is hardly impaired but diagnosis is incomplete, especially before entering and after leaving school. In both mild and severe defect the ratio of incidence to prevalence is roughly 2, or an incidence of 306 severe defectives and 2450 potential mild defectives per 100 000 survivors of infancy. Tarjan *et al* (1973) were led by somewhat different considerations to similar estimates, which, however, reflect the current tendency to diagnose severe mental defect in moribund young children who previously were not committed to institutions for the retarded. In Colchester, for example, no patient

was below age 4 (Penrose, 1938, p. 15). At the opposite extreme of mild defect there is growing pressure for community placement or discharge and for consideration of dull individuals, regardless of their adaptive behaviour, as not retarded. These developments must affect prevalence, incidence, and hospitalization rates (Table II).

Resolution of heterogeneity

While mental retardation is highly heterogeneous, practical considerations force us to limit genetic analysis to a small number of categories. We decided to aim for three aetiological groups, based on a suggestion by Tarjan (1970):

I: Biological

Significant physical or laboratory pathology; cause of retardation apparently intrinsic; diagnosis usually made before school entrance; often abnormal in physical appearance; mortality greater than the general population; retardation moderate or severe; parental IQ and socioeconomic status representative of the general population; low birthweight, postpartum distress, parental consanguinity, or raised parental age in a proportion of cases; parents rarely, sibs sometimes retarded.

II: Medical

Significant physical pathology; a specific, single, extrinsic aetiological agent probably accounts for the condition; diagnosis made not long after environmental insult; abnormal in physical appearance; mortality considerably greater than the general population; retardation severe, either stationary or progressive; parental IQ and socioeconomic status representative of the general population; parents and sibs rarely retarded.

III: Sociofamilial

Currently available biomedical technology cannot show significant physical or laboratory pathology; no single, specific aetiological agent can be made

TABLE II
PREVALENCE AND INCIDENCE IN A POPULATION OF 100 000

Source	Prevalence		Incidence		In Hospital	
	Severe	Mild	Severe	Mild	Severe	Mild
Colchester	153	1225	306	2450	38	37
Tarjan <i>et al</i> (1973)	250	750	500	2500	—	—

Colchester revisited: A genetic study of mental defect

convincingly accountable for the condition; diagnosis usually made after school entrance, and often retracted at young adult age; normal in physical appearance; mortality not much greater than the general population; retardation mild, with the difference between mean parental IQ and that of the patient frequently small; often from the socially, economically, and educationally underprivileged strata of our society.

Biological and medical categories are considered to depend on *megaphenic* effects (i.e. large relative to the standard deviation of IQ so that each case is primarily due to a single, specific etiological agent), while the sociofamilial category is *microphenic* (due to cumulative effects of small unidentifiable causes). Such Platonic entities are indispensable for epidemiology, but they do not allow for dual causation by megaphenic and microphenic effects (for example, greater risk of cerebral infection or birth injury in the lowest social class, or predisposition of a particular genotype to meningitis), and they do not provide an unambiguous classification of individuals. Rather than a subjective classification, it seems better to use a discriminant from less to more megaphenic.

A continuous discriminant can be developed from the following considerations. Penrose (1963) estimated that 99% of mild defect but only 13% of severe defect is part of the Gaussian curve of IQ, which includes but is not limited to microphenic effects. In fact, refinements in diagnostic technique may lead to recognition of megaphenic causes underlying an ostensibly Gaussian distribution, like the three alleles which account for most of the variation of red cell acid phosphatase (Harris, 1975). Segregation analysis of mental defect can do little

with the microphenic component except determine its frequency and recurrence risks, leaving cumulative environmental and genetic factors to be resolved by path analysis in the general population (Rao *et al*, 1974, 1975). Here we shall not attempt to partition this microphenic component, and estimates of heritability will, therefore, include cultural as well as biological inheritance. The goal of the analysis is to delimit the microphenic contribution and isolate components of the megaphenic category caused by genetic mechanisms.

To that end an index of severity is useful, since the proportion of megaphenic cases increases with severity. We used the method of criterion scaling (Beaton, 1969), recording grade of patient from 100 for dull to 400 for idiot and transforming all other variables in terms of the mean value of this criterion (Table III). Variables used by Penrose but not tabulated in his appendix were extracted from the original records by one of us (H.L.B.). Consanguinity was not significantly related to severity and so was omitted. The 14 significant variables were combined by the first principal component of their correlation matrix.

If the assumed categories are valid, the distribution of the severity index should be a mixture of megaphenic and microphenic components. MacLean *et al* (1976) have developed an analysis of admixed distributions which allows for skewness. It was used to define a most sociofamilial group (B_0) corresponding to 9% of cases. This is considerably less than the category without significant physical or metabolic pathology (24%), which must contain megaphenic cases (aminoacidopathies, etc.) not recognized with the diagnostic techniques available at the time of the Colchester study, and in

TABLE III
CRITERION SCALING OF PROBANDS: UNKNOWN AND OTHER DENOTED BY *

Variable	i	Scale (z_i)	σ_i	b_i
Sex	1	$\delta \rightarrow 262, \varphi \rightarrow 245$	8.5	0.076
Grade	2	$M_0 \rightarrow 100, M_1 \rightarrow 200, M_2 \rightarrow 300, M_3 \rightarrow 400$	93.4	0.267
Reliability	3	$1 \rightarrow 261, 2 \rightarrow 236, 3 \rightarrow 238$	11.0	0.297
Grade of father	4	$S \rightarrow 302, N \rightarrow 261, M_0 \rightarrow 236, M_1 \rightarrow 227, M_2 \rightarrow 233, * \rightarrow 208$	15.1	0.352
Grade of mother	5	$S \rightarrow 332, N \rightarrow 265, M_0 \rightarrow 230, M_1 \rightarrow 229, M_2 \rightarrow 261, * \rightarrow 217$	16.9	0.328
Age	6	$394 - 13.5A + 0.35A^2 - 0.0025A^3$	22.0	0.113
Admission	7	$1 \rightarrow 308, 2 \rightarrow 308, 3 \rightarrow 206, 4 \rightarrow 263, 5 \rightarrow 270, 6 \rightarrow 183$	38.7	0.224
Occupation of father	8	$A_1 \rightarrow 306, A_2 \rightarrow 284, A_3 \rightarrow 298, B \rightarrow 277, C \rightarrow 277, D \rightarrow 290, E_1 \rightarrow 256, E_2 \rightarrow 244, * \rightarrow 222$	19.3	0.326
Age of father	9	$230 + 0.83A, * \rightarrow 223$	12.4	0.313
Age of mother	10	$228 + 0.0285A^2, * \rightarrow 229$	12.9	0.215
Rating of home	11	$A \rightarrow 328, B \rightarrow 293, C \rightarrow 257, D \rightarrow 237, E \rightarrow 222, * \rightarrow 242$	20.8	0.332
Size of home	12	$233 + 7.54R - 2.94P, * \rightarrow 248$	10.7	0.236
Illegitimacy	13	$1 \rightarrow 222, 0 \rightarrow 257$	9.3	0.281
AAMD diagnosis	14	$01x \rightarrow 276, 02x \rightarrow 219, 1xx \rightarrow 276, 2xx \rightarrow 275, 3xx \rightarrow 259, 4xx \rightarrow 309, 46x \rightarrow 240, 47x \rightarrow 309, 48x \rightarrow 285, 56x \rightarrow 324, 70x \rightarrow 243, 800 \rightarrow 230, 801 \rightarrow 177, 902 \rightarrow 274, 903 \rightarrow 248$	39.4	0.212

The equation $A \rightarrow N$ signifies that probands in category A are assigned their mean severity N. For example males are scored as 262. The severity index is computed as $Y = \sum_i b_i z_i / \sigma_i$, where σ_i is the standard deviation and b is the eigenvector ($i = 1, \dots, 14$).

TABLE IV
DIAGNOSIS BY AAMD CODE, MEAN INDEX, AND CATEGORY

Code	Diagnosis	Number of Cases	Mean Index	Medical M	Sociofamilial			Mixed B ₃	Biological B ₄
					B ₀	B ₁	B ₂		
01X	Prenatal infection (syphilis)	50	56.5	34	16	—	—	—	—
02X	Postnatal infection	21	57.5	20	1	—	—	—	—
1XX	Birth or postnatal injury	38	57.9	36	2	—	—	—	—
2XX	Metabolic	91	57.4	—	5	15	25	17	29
3XX	Postnatal gross brain disease	17	57.7	—	1	3	6	2	5
4XX	Cranial deformity	82	57.9	—	4	8	18	18	34
46X	Other deformity	83	57.1	—	4	16	29	24	10
47X	Neurological disease	67	58.1	—	1	4	8	31	23
48X	Sensory disease	40	58.1	—	1	3	8	15	13
56X	Down's syndrome	72	58.9	—	1	2	3	17	49
70X	Severe psychopathy	69	57.3	—	2	15	21	19	12
800	Aclicinal	308	56.3	—	42	101	79	59	27
801	Mild psychopathy	132	55.9	—	20	52	40	17	3
902	Certain epilepsy	120	57.3	—	9	16	28	40	27
903	Doubtful epilepsy	90	57.2	—	6	19	22	23	20
Total		1280	57.1	90	115	254	287	282	252

many cases probably not recognizable at the present time.

We also created an index of biological factors, taking as the criterion the value 1 for AAMD codes beginning with 2-7 or 9, considered to be mostly biological, and the value 0 for other AAMD codes (0, 1, 8) with large medical or sociofamilial components. For this purpose the Colchester data were reclassified by the AAMD code (Grossman *et al*, 1973) as used by us for mental retardation in Hawaii (Table IV). The resultant biological index has so high a correlation with the severity index (0.963) that it is impractical to separate biological and medical cases on the basis of variables other than medical history.

The estimate of admixture for the biological index is 0.84, leaving 16% as medical and most sociofamilial. Since the latter were estimated from the severity index to comprise 9% of cases (115 individuals), we conclude that about 90 cases (7%) should be classified as medical. Accordingly we took as the medical category (M) the 90 largest values of the severity index for diagnostic codes beginning with 0 or 1. In the remainder the

115 smallest values of the severity index were taken as most sociofamilial (B₀) and the rest (84%) were divided into quartiles from least to most megaphenic (B₁ to B₄).

The most megaphenic biological group (B₄) contains a substantial number of patients classified by Penrose as without physical or metabolic pathology (Table IV). Most diagnoses contributed few members to B₀, which is drawn largely from aclicinal cases, mild psychopathy, and congenital syphilis. The latter was diagnosed by a variety of criteria, and the possibility of false positives was discussed by Penrose (1938, pp. 38-40), who concluded that 'in some cases defect may be entirely due to the physical disease but in others it merely accentuates an endogenous mental inferiority'. Our analysis, in which 16 of 50 cases are classified as most sociofamilial (B₀), supports this suggestion.

There is a clear relation between aetiological group and severity of retardation, which increases from B₁ to B₄ (Table V). The proportion of sociofamilial cases decreases from mild to severe defect. Most cases classified as extremely sociofamilial (B₀) have one or both parents of unclassified

TABLE V
GRADE BY AETIOLOGY

Grade	IQ Range	Medical	Sociofamilial			Mixed B ₃	Biological B ₄	Total
			B ₀	B ₁	B ₂			
Dull	M ₀ 70-84	9	26	74	42	21	7	179
Simpton	M ₁ 50-69	30	62	125	121	83	27	448
Imbecile	M ₂ 20-49	30	25	48	95	119	116	433
Idiot	M ₃ < 20	21	2	7	29	59	102	220
Total		90	115	254	287	282	252	1280
Incidence/100 000		176	356	804	696	490	234	2756

TABLE VI
MATING TYPE BY AETIOLOGY

Group	No. of Families	No. of Children	N × N	N × I	N × U	N × A	U × U	I × I	A × A	I × U	I × A	A × U
M	86	400	73	7	—	3	—	1	—	—	2	—
B ₀	113	188	8	13	21	15	9	1	—	19	7	20
B ₁	238	1011	54	72	11	24	—	25	15	2	34	1
B ₂	271	1350	162	73	3	18	—	10	3	—	2	—
B ₃	274	1456	251	13	—	7	—	1	—	—	2	—
B ₄	245	1213	244	1	—	—	—	—	—	—	—	—
Total	1227	5618	792	179	35	67	9	38	18	21	47	21

N = normal, I = dull, U = unknown, A = retarded.

intelligence (Table VI). They are often illegitimate and with few if any full sibs. The proportion of cases with one or both parents dull or retarded is high for B₁ and decreases to virtually zero for B₄, the most biological group.

However interpreted, the aetiological categories are more homogeneous than the whole sample. Segregation and consanguinity analysis can tell us more about them.

Segregation analysis of aetiological groups

Table II gives 0.0153 as the frequency of retardation in the general population if incidence is appropriate for severe defect, and prevalence for mild defect. We consider a patient of grade M₀ (dull) or worse and a parent or nonpatient sib of grade M₁ (mildly retarded) or worse as affected. With this definition the ascertainment probability from the distribution of the number of probands per sibship in the 222 different sibships with two or more affected is $\pi = 0.263 \pm 0.029$ (Morton, 1959). This estimate does not vary significantly among aetiological groups, presumably because mortality of severe cases and community placement of mild cases have somewhat similar effects on the ascertainment probability. Therefore, we assume the above value of π for each category. For lack of evidence about the frequency in different segments of the population we also assume the same incidence for each category.

A preliminary analysis of severe mental defect by complex segregation analysis (Morton and MacLean, 1974) agreed closely with the simpler analysis of Dewey *et al* (1965). It also showed that partitioning individuals into 3 classes (normal, intermediate, affected) gives greater power than two classes (normal, affected). Since we are now considering all grades of mental retardation, we define intermediate by the frequency of dull but not retarded individuals in the general population. Table I gives 423 dull individuals among 10 888 second

degree and more remote relatives, a proportion of 0.0388. Classification on the psychometrics rather than conservative judgment might raise this to approximately the frequency in a normal distribution between 1 and 2 standard deviations (IQ 85–70), or 0.136. A lower frequency is appropriate for the Colchester Survey, where the diagnosed frequency of dull sibs was only 0.106 (cp. Reed and Reed, 1965, p. 5). For the displacement of recessive homozygotes we take $t = 3.5$. This is supported by analysis of normal × normal matings which give a slightly smaller likelihood at lower values of homozygous displacement t when heritability H and recessive gene frequency q are estimated simultaneously. At $t = 3.5$ the other estimates are $H = 0.30 \pm 0.03$, $q = 0.03 \pm 0.01$.

Table VII gives the analysis of aetiological categories, separating congenital syphilis and Down's syndrome from the rest of the material and pooling the most sociofamilial categories B₀, B₁, B₂, which on preliminary analysis appeared to be homogeneous in regard to the parameters for heritability H and recessive gene frequency q . This sociofamilial group gives no evidence of recessive genes, but a nonsignificant suggestion of environment common to sibs. Cultural inheritance from parent to child is confounded with heritability, the estimate of which is 0.833 ± 0.026 . While the hypotheses that heritability in the sociofamilial group is entirely genetic or entirely cultural cannot be disproven on these data, such limiting cases seem unlikely since IQ is determined by both genetic and cultural inheritance (Rao *et al*, 1974).

The most megaphenic group (B₄) gives highly significant evidence for recessive genes ($q = 0.048 \pm 0.009$), but no suggestion of heterozygous expression or polygenic heritability. Since there is no significant role for other familial mechanisms, most non-recessive cases are sporadic, with negligible recurrence risks.

The intermediate group (B₃) seems to be a mixture of sociofamilial and megaphenic cases, with

TABLE VII
SEGREGATION ANALYSIS OF MENTAL DEFECT AS A TRICHOTOMY
H = heritability; B = effect of environment common to sibs; q = recessive gene frequency.

Group	Assuming $q = B = 0$		Estimating H and q Simultaneously at $B = 0$		
	Heritability $\pm$ SE	χ^2	Heritability $\pm$ SE	Gene frequency $\pm$ SE	χ^2
Down's syndrome	0.082 $\pm$ 0.084	159.51	0.011 $\pm$ 0.080	0.035 $\pm$ 0.018	157.64
Congenital syphilis	0.700 $\pm$ 0.106	84.38	0.701 $\pm$ 0.109	0.018 $\pm$ 0.072	84.35
Other medical	0.400 $\pm$ 0.110	137.36	0.400 $\pm$ 0.110	0	137.36
Remainder:					
Sociofamilial ($B_0 + B_1 + B_2$)	0.833 $\pm$ 0.026	2472.13	0.831 $\pm$ 0.026	0.018 $\pm$ 0.011	2471.25
Mixed (B_3)	0.603 $\pm$ 0.061	995.32	0.585 $\pm$ 0.062	0.018 $\pm$ 0.014	994.78
Biological (B_4)	0.198 $\pm$ 0.044	392.77	0	0.048 $\pm$ 0.009	369.81
Total	0.748 $\pm$ 0.025	4314.01	0.747 $\pm$ 0.025	0.016 $\pm$ 0.007	4312.23

Group	Assuming Gene Frequency $q = 0$			Assuming $q = H = B = 0$
	Heritability $\pm$ SE	Common environment $\pm$ SE	χ^2	χ^2
Down's syndrome	0.082 $\pm$ 0.084	0	159.51	160.30
Congenital syphilis	0.439 $\pm$ 0.270	0.179 $\pm$ 0.137	81.46	123.96
Other medical	0.400 $\pm$ 0.110	0	137.36	154.00
Remainder:				
Sociofamilial ($B_0 + B_1 + B_2$)	0.801 $\pm$ 0.037	0.030 $\pm$ 0.019	2468.82	3994.09
Mixed (B_3)	0.603 $\pm$ 0.061	0	995.32	1176.50
Biological (B_4)	0.198 $\pm$ 0.044	0	392.77	406.47
Total	0.748 $\pm$ 0.025	0	4314.01	6015.32

Frequency of affected (A) = 0.0153, frequency of intermediates (I) = 0.0388, dominance (d) = 0, displacement (t) = 3.5, and ascertainment probability (π) = 0.263. The columns indicated as χ^2 are computed as $-2\ln L + \text{const.}$ For example, the test of $q = 0$ in group B_4 is $\chi^2 = 392.77 - 369.71 = 23.06$.

intermediate heritability (0.603 $\pm$ 0.061). There is no significant evidence of recessive genes, but the standard error of the gene frequency is large. If we assume admixture in proportion x of families with megaphenic cases indistinguishable from B_4 ($H = 0$, $q = 0.048$), the remaining fraction $1 - x$ of families having microphenic cases ($H = 0.833$, $q = 0$), the

likelihood is appreciably increased ($\chi^2 = 970.55$). Genetic counselling in this category should, therefore, be made on the hypothesis of admixture rather than from the intermediate values of H and q estimated directly. The maximum likelihood estimate of the frequency of megaphenic cases in B_3 is 0.59 $\pm$ 0.09.

TABLE VIII
PARENTAL CONSANGUINITY BY ETIOLOGY

Aetiology	Inbreeding Coefficient F							Total	Mean F $\times 10^5$
	0	1/128	1/64	1/32	1/16	1/8	1/4		
Medical	90	—	—	—	—	—	—	90	0
Sociofamilial ($B_0 + B_1 + B_2$)	634	4	—	3	11	2	2	656	238
Mixed (B_3)	267	1	—	3	9	2	—	282	324
Biological (B_4)	244	—	2	1	4	1	—	252	174
Total	1235	5	2	7	24	5	2	1280	228

NORMAL $\times$ NORMAL MATINGS

Sociofamilial ($B_0 + B_1 + B_2$)	224	1	—	—	2	—	—	227	59
Mixed (B_3)	244	1	—	2	7	1	—	255	248
Biological (B_4)	243	—	2	1	4	1	—	251	174
Total	711	2	2	3	13	2	—	733	164

Bell (1940), hospitals	48 926	15	53	2	299	3	0	49 298	41
Böök (1957), retarded	81	—	—	—	7	—	—	88	
Böök (1957), normal	6259	—	—	—	188	—	—	6377	

Inbreeding

Table VIII gives clear evidence that inbreeding, and especially incest, is more frequent for dull or retarded individuals than for normals. Consanguinity is conspicuously raised among the parents of the sociofamilial group, but is not appreciably more frequent than in the control population when only normal $\times$ normal matings are considered.

For $B_3 + B_4$, on the other hand, the striking association with parental inbreeding persists in normal $\times$ normal matings. If we take 0.0128 as the incidence in these matings, the method of Morton (1960) gives

$$A = \text{panmictic load} = 0.0124$$

$$B = \text{inbred load} = 0.893 \pm 0.263$$

For Swedish data (Böök, 1957) the corresponding inbred load is

$$B = 0.716 \pm 0.349$$

and so the pooled estimate is $\hat{B} = 0.83 \pm 0.21$. Since this is about four times the load for severe mental defect, it confirms the suggestion of Dewey *et al* (1965) that 'some rare recessive genes act to produce mild retardation, though they are difficult to discriminate because of overlap with the polygenes and environmental factors causing the lower part of the continuous, nearly normal distribution of intelligence'.

The incidence of homozygotes in the general population served by Colchester has been estimated as 0.048^2 (Table VII). Deducting B_α , the expressed load resulting from an inbreeding coefficient α , we have

$$A = 1964 \times 10^{-6}$$

as the panmictic incidence of mental retardation caused by rare autosomal recessive genes, or six times the estimate of Dewey *et al* (1965, p. 254) for severe mental defect. The mean gene frequency per contributory locus is $Q \leq A/B = 0.0024$, and the number of contributory loci is $k \geq B^2/A = 351$. These estimates could easily be in error by a factor of 2, but they indicate that the number of loci is large and the individual gene frequencies small. Assuming a mutation-selection equilibrium, the mutation rate per gamete is about $0.01B = 0.008$ and the mutation rate per locus is $\mu \leq 0.01Q = 2.4 \times 10^{-5}$, in reasonable agreement with other rare recessive genes. The possibility that heterozygote advantage plays an important role in maintaining this load is remote (Dewey *et al*, 1965). In that event, the mutation rates would be smaller but the estimates of gene frequency and number of loci would not be affected.

Sex ratio

If sex linkage accounts for part of the excess of retarded males, as has been suggested by Lehrke (1974), there should be an excess of families with only affected males corresponding to the distribution

$$P(r=0) = h + (1-h)(1-p)^s$$

$$P(r>0) = (1-h)\binom{s}{r}p^r(1-p)^{s-r}$$

where h is the proportion unable to segregate affected girls, p is the probability of girls among affected, r is the number of affected girls, and $s-r$ is the number of affected boys (Morton, 1959). For the 1227 Colchester families there is no doubt that males are in excess among affected ($\chi^2 = 14.6$), but the proportion of cases caused by sex linkage is in doubt (Table IX). The maximum estimate of the proportion of sex-linked families h for $p \leq 0.5$ is at the limit $p = 0.5$, with $h = 0.089 \pm 0.023$, and the likelihood is virtually the same as for h and p estimated simultaneously. The hypothesis that the sex ratio $p = 0.5$ is admissible, and among normal sibs the proportion of females is 0.494 ± 0.008 , as in the general population. However, it is unlikely that boys and girls are equally frequent among cases not due to major sex-linked genes, since retarded boys are more difficult to manage and polygenic variability must be greater for the heteromorphic sex. If polygenes are distributed between autosomes and the X chromosome in proportion to their length and if the heritability of IQ in the general population is 0.5 (Rao *et al*, 1974), then the sex ratio of affected would be

$$p = \frac{0.494}{0.494 + 0.506e^{0.025}} = 0.488,$$

at which value the estimate of h is 0.070 ± 0.024 . This is not significantly different from the estimate of 0.040 ± 0.046 when p and h are iterated simultaneously. We tentatively conclude that about 5% of cases are the result of sex linkage.

Discussion

This analysis depends on the basic soundness of the Colchester survey, which includes data essential

TABLE IX
THE SEX OF AFFECTED

Hypothesis	χ^2
Equal sex ratio, no sex-linkage ($p = 0.5, h = 0$)	37.13
Equal sex ratio, sex-linkage ($p = 0.5, \hat{h} = 0.089 \pm 0.023$)	23.20
Unequal sex ratio, no sex-linkage ($\hat{p} = 0.452 \pm 0.013, h = 0$)	22.52
Unequal sex ratio, sex-linkage	
($\hat{h} = 0.470 \pm 0.025, \hat{p} = 0.040 \pm 0.046$)	21.95
Theoretical sex ratio, sex-linkage	
($p = 0.488, \hat{h} = 0.070 \pm 0.024$)	22.43

for resolution of heterogeneity and complex segregation analysis. However, with the wisdom of 40 years' hindsight some refinements can be suggested. Replacement of intellectual grades by psychometrics on all probands, sibs, and parents would allow more powerful segregation analysis and more reliable estimates of incidence. Greater symptomatic detail (and of course application of the cytogenetic, biochemical, and other diagnostic advances of the last generation) would give better resolution of aetiological heterogeneity.

These data provide no critical distinction between cultural inheritance and polygenes in the large proportion of sociofamilial cases. The distinction, while interesting, is irrelevant to recurrence risks in first-degree relatives, which can be predicted from the values of segregation analysis (Maclean *et al*, 1976).

On the evidence there is clear separation of the extreme groups of sociofamilial retardation ($B_0 + B_1 + B_2$) from megaphenic defect (B_4) after exclusion of recognized exogenous types (M). The distribution of cases is estimated in Table X for the Colchester survey and the general population. The majority are sociofamilial, only about 7% are the result of recognized exogenous factors, and perhaps a third are biological. Since premature mortality reduces prevalence of severe defect by about one-half and incomplete diagnosis has roughly the same effect on prevalence of mild defect, the proportions are much the same for incidence and prevalence. The medical category would presumably be more frequent under the modern convention of classifying moribund young children with multiple

anomalies as severely retarded, whereas they were excluded in the Colchester survey.

The incidence of homozygotes for rare autosomal recessive genes causing mental retardation was estimated above as $0.048^2 = 0.0023$. The incidence of biological cases ($B_4 + 0.59 B_3$) is 0.0052. Apparently autosomal recessives account for less than half of the biological category, the remainder being due to other mechanisms such as sex-linkage, recent dominant mutations (both genic and chromosomal), aneuploidy, unrecognized medical factors, and perhaps more complicated causes.

With respect to sex-linkage our results are inconclusive, since the distribution of affected males and females is consistent with a negligible or a considerable proportion of sex-linked cases, depending on the sex-ratio of affected in other cases (p). If 5% of genetic cases are the result of sex-linkage, as might be expected if the loci which produce mental retardation are randomly distributed among chromosomes (Paris Conference, 1971), then the incidence of sex-linked recessive types of mental defect would be less than 0.0003. However, this neglects the reduction in incidence of homozygotes for autosomal recessives as inbreeding declined during the last century. If the mean gene frequency is only 0.0024 per contributing locus (as estimated), then most homozygotes at a mutation-selection equilibrium were due to identity by descent. If so, at present reduced levels of inbreeding sex-linked recessives might make up several per cent of all cases, without any disproportionate concentration on the X chromosome of genes for mental retardation. At 5% of all cases the number of Colchester probands

TABLE X
ESTIMATED DISTRIBUTION OF CASES AND INCIDENCE

Source		Medical (M)	Sociofamilial ($B_0 + B_1 + B_2 + 0.41B_3$)	Biological ($B_4 + 0.59B_3$)	Total
Colchester survey (Table IV)	N	90	772	418	1280
	%	7	60	33	100
Incidence/100 000 (Table V)	N	176	2056	524	2756
	%	6	75	19	100

Prevalence $\pm$ incidence/2.

TABLE XI
ESTIMATED DISTRIBUTION OF BIOLOGICAL CASES AND INCIDENCE

Source		Sex-linked Recessive	Autosomal Recessive	Dominant	Down's Syndrome	Other Chromosomal	Total
Colchester survey (1280 cases)	No.	64	169	30	72	24	359
	%	18	47	8	20	7	100
Incidence/100 000	No.	87	230	40	138	46	541
	%	16	43	7	26	8	100

Colchester revisited: A genetic study of mental defect

would be 64. If these fall into the B₃ and B₄ categories they might from Table V have an incidence of 2(64)(0.00724)/534, or 174 per 100 000 male births. If fertility of these hemizygotes is negligible, but heterozygous females are normal, and if mutation is equally frequent in egg and sperm, the mutation rate per X chromosome would be 58×10^{-5} . We estimate the number of contributory loci per autosomal genome to be at least 351. If the X chromosome carries a proportionate number of recessive genes there would be at least 18 X-linked loci that could cause mental defect and the mutation rate per locus would be at most 3.2×10^{-5} , a typical value.

Although any attempt to partition the biological category must be approximate, it seems worth while to make the attempt (Table XI). We have just estimated the sex-linked component and the incidence of autosomal recessives. If the numbers of cases are proportional to the incidences, as would be expected if sex-linked and autosomal recessives had on average similar degrees of severity, there would be about 169 autosomal recessive cases in the Colchester survey, of whom 22 have normal but consanguineous parents (Table VIII).

Among the coded diagnoses are 6 cases of neurofibromatosis, 5 of epiloia, 1 of Huntington's chorea, and 21 of acrocephaly (including acrocephalosyndactyly). From Bergsma (1974) the incidence of dominant types of mental retardation is about 40 per 100 000. On the same assumptions as for recessives, this would give about 30 dominant cases in the Colchester survey.

Newton *et al* (1972) conducted a cytogenetic survey of 1255 patients over 15 years of age in a hospital for the mentally subnormal. There were 104 cases of Down's syndrome, in rough agreement with the 72 cases of the Colchester survey in samples of virtually the same size. The incidence of Down's syndrome is about 138 per 100 000 births (Penrose, 1963, p. 208). There were 24 other chromosomal abnormalities in the cytogenetic survey, compared with 422 per 100 000 in the general population (Jacobs *et al*, 1974), of which about 1/10 cause mental retardation (P. A. Jacobs, 1976, personal communication).

Adding the genetic categories together, we observe that their frequency is approximately B₄ + 0.59 B₃, as expected. Most of the biological group is the result of simple mechanisms, with little if any residuum where unrecognized exogenous or complex causes could be involved.

REFERENCES

- Beaton, A. E. (1969). Scaling criterion of questionnaire items. *Socio-Economic Planning Sciences*, **2**, 355-362.
- Bell, J. (1940). A determination of the consanguinity rate in the general hospital population in England and Wales. *Annals of Eugenics*, **10**, 370-391.
- Bergsma, D. (1974). *Birth Defects: Atlas and Compendium*. Williams and Wilkins, Baltimore (The National Foundation).
- Böök, J. A. (1957). Genetical investigations in a north Swedish population: the offspring of first-cousin marriages. *Annals of Human Genetics*, **21**, 191-221.
- Crow, J. F. (1968). Some analysis of hidden variability in Drosophila populations. In *Population Biology and Evolution*, pp. 71-86. Ed. by R. C. Lewontin. Syracuse University Press, New York.
- Dewey, W. J., Barrai, I., Morton, N. E., and Mi, M. P. (1965). Recessive genes in severe mental defect. *American Journal of Human Genetics*, **17**, 237-256.
- Grossman, H. J., Warren, S. A., Begab, M. J., Eyman, R., Nihira, K., and O'Connor, G. (1973). *Manual on Terminology and Classification in Mental Retardation (Revised)*. American Association of Mental Deficiency, Special Publications Series 2.
- Harris, H. (1975). *The Principles of Human Biochemical Genetics*. Elsevier, New York.
- Jacobs, P. A., Melville, M., Ratcliffe, S., Keay, A. J., and Syme, J. (1974). A cytogenetic survey of 11,680 newborn infants. *Annals of Human Genetics*, **37**, 359-376.
- Lehrke, R. G. (1974). *X-linked Mental Retardation and Verbal Disability*. Birth Defects: Original Article Series, 10, No. 1. Intercontinental Medical Book Corporation, New York.
- MacLean, C. J., Morton, N. E., Elston, R. C., and Yee, S. (1976). Skewness in commingled distributions. *Biometrics*. In the press.
- Morton, N. E. (1959). Genetic tests under incomplete ascertainment. *American Journal of Human Genetics*, **11**, 1-16.
- Morton, N. E. (1960). The mutational load due to detrimental genes in man. *American Journal of Human Genetics*, **12**, 348-364.
- Morton, N. E. (1974). Analysis of family resemblance. I. Introduction. *American Journal of Human Genetics*, **26**, 318-330.
- Morton, N. E. and MacLean, C. J. (1974). Analysis of family resemblance. III. Complex segregation of quantitative traits. *American Journal of Human Genetics*, **26**, 489-503.
- Newton, M. S., Cunningham, C., Jacobs, P. A., Price, W. H., and Fraser, I. A. (1972). Chromosome survey of a hospital for the mentally subnormal. Part 2: Autosomal abnormalities. *Clinical Genetics*, **3**, 226-248.
- Paris Conference (1971). Standardization in human cytogenetics. *Birth Defects: Original Article Series*, **8**, No. 7. The National Foundation, New York, 1972.
- Penrose, L. S. (1938). A clinical and genetic study of 1,280 cases of mental defect. *Medical Research Council. Special Report Series*, No. 229.
- Penrose, L. S. (1963). *The Biology of Mental Defect*. Sidgwick and Jackson, London.
- Rao, D. C., MacLean, C. J., Morton, N. E., and Yee, S. (1975). Analysis of family resemblance. V. Height and weight in north-eastern Brazil. *American Journal of Human Genetics*, **27**, 509-520.
- Rao, D. C., Morton, N. E., and Yee, S. (1974). Analysis of family resemblance. II. A linear model for familial correlation. *American Journal of Human Genetics*, **26**, 331-359.
- Reed, E. W. and Reed, S. L. (1965). *Mental Retardation: A Family Study*. W. B. Saunders, Philadelphia.
- Tarjan, G. (1970). Some thoughts on socio-cultural retardation. In *Socio-cultural Aspects of Mental Retardation*. Ed. by H. C. Haywood. Appleton-Century-Crafts, New York.
- Tarjan, G., Wright, S. W., Eyman, R. K., and Keeran, C. V. (1973). Natural history of mental retardation: some aspects of epidemiology. *American Journal of Mental Deficiency*, **77**, 369-379.