J. Genet. Vol. 71, No. 3, December 1992, pp. 89–103.

Statistical analysis of family data on complex disorders in man*

PARTHA P. MAJUMDER and SWAPAN K. NATH
Anthropometry & Human Genetics Unit, Indian Statistical Institute, Calcutta 700 035, India

MS received 6 November 1992

Abstract. A genetic model is discussed for recessively inherited disorders that do not follow a single-locus Mendelian pattern of inheritance. Further complexity arising from variable age of onset is also discussed. Methods of statistical analysis of family data using the likelihood principle are described for such complex disorders. The methods are exemplified using data on families of prelingual deafness and vitiligo.

Keywords. Segregation analysis; autosomal recessive; multiple loci; variable onset age; prelingual deafness; vitiligo.

1. Sources of complexity of a disorder

In comparison with a genetic disorder which has no environmental contribution, expresses itself at birth, and is determined by a completely penetrant allele at a single autosomal diallelic locus, the genetics of many human disorders is more complex. The complexity can arise in a variety of ways. Although no claim is made that the following list is exhaustive, some of the more common causes of complexity of a disorder are:

(i) Variable age of onset—all individuals with the appropriate genotype do not manifest the disorder either at birth or at the same age later in life;
(ii) Incomplete penetrance—some individuals with the appropriate genotype manifest the disorder while some others do not;
(iii) Phenotypic heterogeneity—all individuals of the same genotype do not manifest the same phenotype;
(iv) Allelic/genetic heterogeneity—different alleles, either at the same locus or at different loci, give rise to the same phenotype;
(v) Multiple loci—disorder is determined by the action of genes at more than one locus;
(vi) Environmental influence—environment interacts with genotype in the manifestation of the disorder.

The causes listed above are not mutually exclusive. For example, a disorder may be determined jointly by the action of genes and environment.

2. Scope of the present study

In this paper we consider two sources of complexity of a disorder: (i) complexity arising due to involvement of multiple loci, and (ii) additional complexity arising due

*Based on a talk given at the Haldane Centenary Symposium held on 6 November 1992 at Ahmedabad as part of the 58th Annual Meeting of the Indian Academy of Sciences.

to variability of age of onset. Further, we consider only autosomal recessive disorders. The primary objective will be to present methods of statistical analysis, based on the likelihood principle, of data on families with a view to understanding the genetics of such complex disorders.

3. A model

This model was first explicitly considered by Li (1953) and was later developed more extensively by him in 1987 (Li 1987). Li's reconsideration in 1987 was prompted by David Lykken's Presidential Address to the Society for Psychophysiological Research (Lykken 1982), in which Lykken drew attention to some behavioural traits (e.g. mathematical ability) which tend not to run in families, but appear as isolated cases. (Carl Frederick Gauss or Srinivasa Ramanujan may be cited as examples.) Patterns of inheritance of such traits are drastically different from the single-locus inheritance pattern. While such traits may appear to be non-genetic because of their lack of familial aggregation, they may really be genetic. Li (1987) proposed a concrete genetic model for such traits. The model states that such a trait or disorder is determined by a particular combination of genes of a number of loci, thus forming a gene "configuration" or "constellation". Absence of any one gene in this configuration destroys the manifestation of the trait or disorder. More concretely, suppose we consider a large number (say k) of epistatically interacting autosomal diallelic loci (A_i, a_i), $i = 1, 2, \ldots, k$, the model states that the gene configuration responsible for manifestation of the trait or disorder is $(a_1a_1a_2a_2 \ldots a_ka_k)$; individuals of the remaining $3^k - 1$ genotypes do not manifest the trait.

4. Prevalence in population

Suppose q_i denotes the frequency of the allele a_i $(i = 1, 2, \ldots, k)$ in a population. If the population is random-mating, then the prevalence (δ) of the disorder in the population will be

$$\delta = \prod_{i=1}^{k} q_i^2. \tag{1}$$

If $q_i = q$ (for all $i = 1, 2, \ldots, k$), then

$$\delta = q^{2k}. \tag{2}$$

Figure 1 provides the $\log_e$ prevalence rates as the number of loci increases, for various values of q. It is seen that prevalence sharply decreases with increase in the number of loci for a fixed value of the gene frequency. Similar decrease in prevalence is also seen with decrease in gene frequency for a fixed number of loci.

Suppose, instead of practising random mating, the population practises inbreeding. Suppose the population inbreeding coefficient is F. Then, at the ith locus $(i = 1, 2, \ldots, k)$, the frequency of a_ia_i homozygotes will be $q_iF + q_i^2(1 - F)$. If $q_i = q$ $(i = 1, 2, \ldots, k)$, then

$$\delta = [qF + q^2(1 - F)]^k. \tag{3}$$

Figure 1 also provides a comparison of prevalence rates in a random-mating $(F = 0)$

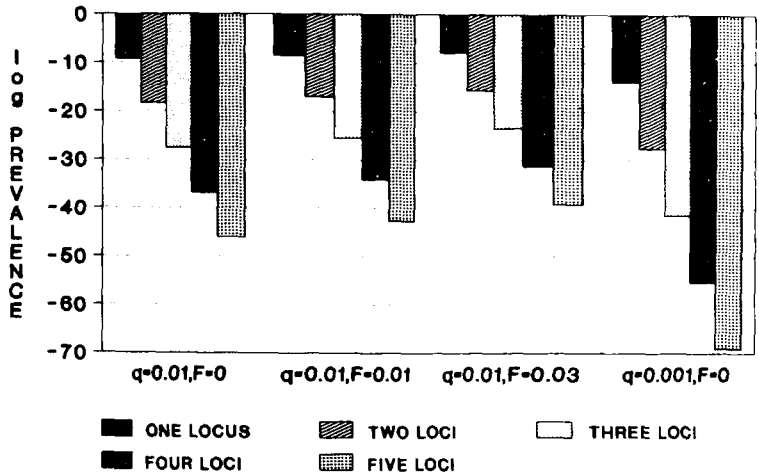

Figure 1. Prevalence of a recessive disorder in random-mating and inbred populations for different numbers of loci.

Table 1. Mating types, their frequencies and segregation probabilities when families are ascertained through an affected offspring.

Mating type	No. of genotypic matings	Frequency*	P(offspring affected given parental mating type)
Normal × Normal	$3^k(3^k-1)/2$	$(1-Q^2)^2$	$S_2 = S_1^2$
Normal × Affected	3^k-1	$2Q^2(1-Q^2)$	$S_1 = Q/(1+Q)$
Affected × Affected	1	Q^4	$S_0 = 1$
Total	$3^k(3^k+1)/2$	1	—

*$Q = \prod_{i=1}^{k} q_i$

versus an inbred ($F > 0$) population. As is expected, prevalence increases with increasing levels of inbreeding.

5. Families in population

Since there are only two phenotypes (normal and affected), there will be three different types of matings in the population: normal × normal (N × N), normal × affected (N × A), and affected × affected (A × A). However, because individuals of the normal phenotype comprise $3^k - 1$ genotypes, N × N matings comprise a vast number of different genotypic matings. This information, adapted from Li (1987), is provided in table 1.

6. Sampled families

While the above results are a description of characteristics of the disorder among individuals in a population or among families drawn at random from the population, the mode of inheritance of such a disorder cannot be effectively studied—especially if its prevalence in the population is low—by sampling individuals or families at random from the population. For such studies, families need to be sampled non-randomly, because the vast majority of randomly drawn families will not even be capable of producing an affected child. The standard methods of sampling families non-randomly are (i) to ascertain families through an affected parent, and (ii) to ascertain them through an affected child. When a family is sampled through an affected parent, no correction of any kind is necessary to take into account the effect of non-random sampling. The observed proportion of affected offspring in the sampled N × A families is compared with the expected proportion S_1 given in table 1.

When families are ascertained through an affected child, the observed proportion of affected offspring is grossly upwardly biased because not only are these families capable of producing an affected child, but only those families that have actually produced an affected child are being studied. Later we discuss some methods of correcting for this ascertainment bias in the analysis of family data for determining the mode of inheritance of a complex disorder. Before that, we discuss some more characteristics of the model under consideration.

7. Proportions of different types of families under biased ascertainment

When families are sampled through an affected offspring, the parental matings may be N × N, N × A, or A × A, as stated earlier. In a sample of families ascertained through an affected offspring, the proportions of the three mating types are functions of the number of loci involved in the determination of the phenotype. Therefore, these proportions may yield clues to the number of loci involved.

Since ascertainment is through an affected child, the father has to be heterozygous at i loci ($i = 0, 1, 2, \ldots, k$) and recessive homozygous at $(k - i)$ loci, and the mother heterozygous at j loci ($j = 0, 1, 2, \ldots, k$) and recessive homozygous at $(k - j)$ loci; k denotes the number of loci involved in the determination of the phenotype.

If $i = 0$ and $j = 0$, both parents are affected;

if $i \neq 0$ and $j \neq 0$, both parents are normal;

if $i \neq 0$ and $j = 0$, the father is normal and the mother is affected; and

if $i = 0$ and $j \neq 0$, the father is affected and the mother is normal.

Now, the total mating frequency given that the family is ascertained through an affected child is:

$$T = \sum_{i=0}^{k} \sum_{j=0}^{k} \binom{k}{i} \binom{k}{j} H^{i+j} R^{2k-i-j}, \tag{4}$$

where $H = 2q(1 - q)$ is heterozygosity and $R = q^2$ is recessive homozygosity. (Although H and R are assumed to be the same at all loci for algebraic simplicity, generalization

to the case where allele frequencies are different at the various loci is easy.) Also,

$$\left.\begin{aligned} &p_{1\,1}, \text{ Prob (both parents are affected)} = R^{2k}, \\ &p_{0\,0}, \text{Prob (both parents are normal)} \\ &\quad = \sum_{i=1}^{k}\sum_{j=1}^{k}\binom{k}{i}\binom{k}{j}H^{i+j}R^{2k-i-j}, \\ &\text{and} \\ &p_{1\,0}, \text{Prob (one parent is affected and the other normal)} \\ &\quad = R^k\left[\sum_{j=1}^{k}\binom{k}{j}H^jR^{k-j} + \sum_{i=1}^{k}\binom{k}{i}H^iR^{k-i}\right] \\ &\quad = 2R^k\left[\sum_{i=1}^{k}\binom{k}{i}H^iR^{k-i}\right]. \end{aligned}\right\} \quad (5)$$

Hence, in the set of ascertained families, the probabilities of various mating types are:

$$\left.\begin{aligned} &\text{Prob(N} \times \text{N)} = p_{0\,0}/T, \\ &\text{Prob(N} \times \text{A)} = p_{1\,0}/T, \\ &\text{and} \\ &\text{Prob(A} \times \text{A)} = p_{1\,1}/T. \end{aligned}\right\} \quad (6)$$

The percentage distributions of the different types of matings in a sample of families ascertained through an affected offspring are given in figure 2 for different values of the prevalence rate and the number of loci. As is expected, the proportion of N × N

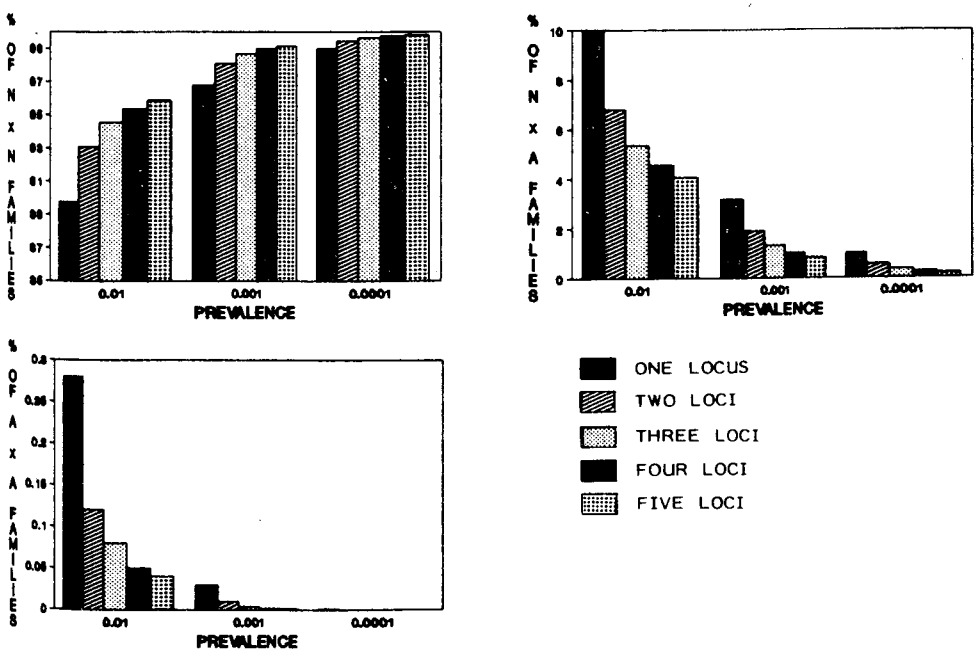

Figure 2. Percentage distributions of mating types in families ascertained through an affected offspring for different levels of prevalence and numbers of loci.

matings increases with an increase in the number of loci for a fixed prevalence rate, and the proportions of N × A and A × A matings correspondingly decrease. The trend is similar if the prevalence rate decreases for a fixed number of loci. Unfortunately, however, for a fixed prevalence rate, the proportions of the three mating types do not change drastically with increase in the number of loci. This means that it is virtually impossible to determine the number of loci using data on proportions of different mating types. To exemplify, using the proportions of N × N families, one would need 545 families to distinguish a two-locus model when the disease frequency is 0·01; when the disease frequency is 0·001, the number of families that would be necessary is 1142.

8. Likelihood of an N × N family ascertained through an affected offspring

To derive the likelihood of an N × N family ascertained through an affected offspring, we first note that each of the normal parents must either be heterozygous at each of the k loci or recessive homozygous at all k loci except at least one. This is because, to produce an affected (aabbcc...) offspring, each parent must be capable of transmitting an abc... gamete, and the reason why neither parent can be recessive homozygous at all the k loci is that each parent is known to be phenotypically normal. Hence, for any such family in which the father is heterozygous at i loci and the mother is heterozygous at j loci, the mating frequency, M_{ij}, is

$$M_{ij} = \frac{\binom{k}{i}\binom{k}{j} H^{i+j} R^{2k-i-j}}{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{k}{i}\binom{k}{j} H^{i+j} R^{2k-i-j}}. \tag{7}$$

The probability θ_{ij} that this family produces an affected offspring is

$$\theta_{ij} = \frac{1}{2^{i+j}}. \tag{8}$$

Now, the probability, α_r, that a family with r affected offspring will have at least one proband is

$$\alpha_r = 1 - (1-\pi)^r, \tag{9}$$

where π denotes the conditional probability that an offspring is a proband given that he or she is affected, which we assume is independent of the parental mating type.

The probability, $\tau_{sr}^{(ij)}$ ($r = 1, 2, 3, \ldots, s$), that a family of the ijth type (that is, in which the father is heterozygous at i loci and the mother is heterozygous at j loci) of size s will have r affected children is

$$\tau_{sr}^{(ij)} = \binom{s}{r} \theta_{ij}^r (1-\theta_{ij})^{s-r}. \tag{10}$$

Therefore, the probability, $\phi_{sr}^{(ij)}$, that a family of ijth type of size s will have r affected children and will be ascertained is

$$\phi_{sr}^{(ij)} = \tau_{sr}^{(ij)} \alpha_r; \quad r = 1, 2, 3, \ldots, s. \tag{11}$$

Hence, the probability, $\Phi_s^{(ij)}$, of a family of *ij*th type of size s having at least one affected child and being ascertained is

$$\Phi_s^{(ij)} = \sum_{r=1}^{s} \phi_{sr}^{(ij)} = 1 - (1 - \pi\theta_{ij})^s. \tag{12}$$

It therefore follows that the probability that a family of size s will have at least one affected child and will be ascertained is

$$\sum_{i=1}^{k}\sum_{j=1}^{k} M_{ij}\Phi_s^{(ij)} = \sum_{i=1}^{k}\sum_{j=1}^{k} M_{ij}[1 - (1 - \pi\theta_{ij})^s]. \tag{13}$$

The likelihood, L, of an ascertained family of size s having r affected offspring then is

$$L = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} M_{ij}\phi_{sr}^{(ij)}}{\sum_{i=1}^{k}\sum_{j=1}^{k} M_{ij}\Phi_s^{(ij)}}. \tag{14}$$

Under single ascertainment, $\alpha_r \approx r\pi$. Hence, L reduces to:

$$L = \binom{s}{r}\binom{r}{s}\frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{k}{i}\binom{k}{j}2^{(i+j)(1-s)}(2^{i+j}-1)^{s-r}p^{i+j}q^{4k-i-j}}{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{k}{i}\binom{k}{j}p^{i+j}q^{4k-i-j}}, \tag{15}$$

where $p = 1 - q$.

Although the above equation looks formidable, considerable simplification is obtained because several mating types have the same values of i and j, and consequently the same value for θ_{ij}. This is exemplified in table 2.

While the above likelihood equation has been derived for unrelated parents, extension to the case where the parents are related is straightforward. The equation remains valid; the only modification that is necessary is in the derivation of the mating probabilities. These probabilities can be derived using the I–T–O method (Li and

Table 2. Parental genotypic mating classes, segregation probabilities and mating frequencies for the two-locus recessive model.

Class	Genotypic mating: Father × Mother	No. of heterozygous loci: Father (i)	No. of heterozygous loci: Mother (j)	Probability of affected offspring (θ_{ij})	Unconditional mating freq. for class	Conditional mating freq. for class
1	AaBb × AaBb	2	2	1/16	$16p^4q^4$	p^2
2	AaBb × aaBb	2	1	1/8	$32p^3q^5$	$2pq$
	aaBb × AaBb	1	2	1/8		
	AaBb × Aabb	2	1	1/8		
	Aabb × AaBb	1	2	1/8		
3	aaBb × aaBb	1	1	1/4	$16p^2q^6$	q^2
	aaBb × Aabb	1	1	1/4		
	Aabb × aaBb	1	1	1/4		
	Aabb × Aabb	1	1	1/4		

Sacks 1954). For the two-locus model, the unconditional mating probabilities as given in table 2 for unrelated parents change to:

when parents are an uncle–niece pair:

$$\left.\begin{aligned} &\text{class 1, } p^2q^2(1\ 2 + 2pq)^2; \\ &\text{class 2, } p^2q^3(1 + 4pq)(1 + 2q); \end{aligned}\right\} \tag{16}$$

and

$$\text{class 3, } 2pq^4\ \{[(1 + q)(1 + 4q) + p(1 + 2q)^2]/4\};$$

when parents are a pair of first cousins:

$$\left.\begin{aligned} &\text{class 1, } 1\ 16 + 3p^3q^3(1/2 + 3pq); \\ &\text{class 2, } p^2q^3[1/4 + 3q/2 + 3pq(1 + 6q)]; \end{aligned}\right\} \tag{17}$$

and

$$\text{class 3, } pq^4[(1 + q) + 12pq(1 + 3q)]/4.$$

9. Distribution of the number of affected children in N × N families under single ascertainment

Using equation 15, the probability distribution of the number of affected children in families where both parents are normal can easily be worked out for the case of single ascertainment. These distributions are presented in figure 3 for various values

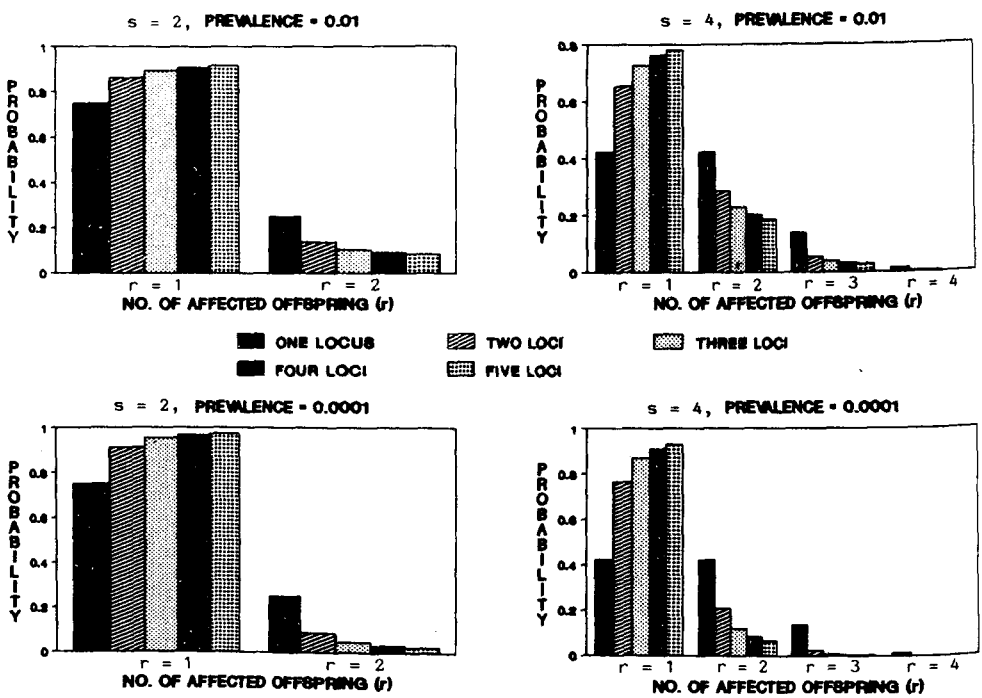

Figure 3. Probability distributions of the number of affected children in families ascertained through an affected child for different sibship sizes and levels of prevalence.

of s (sibship size), δ (prevalence) and k (number of loci). It is seen that for fixed values of s and δ, the probability that a family contains only one affected child monotonically increases with increase in the number of loci. Correspondingly, there is a monotonic decrease in the probability of a family containing a larger number of affected children. While for the one-locus model, the probability distribution of the number of affected children depends only on sibship size and not on prevalence, for multi-locus models this probability distribution also depends on prevalence. As is expected, for multi-locus models, the probability that a family contains a single affected child increases with decrease in prevalence.

10. Sample size considerations

From the probability distribution, the mean and variance of the number of affected children in N × N families can easily be worked out. These statistics can then be used to answer a question of relevance: what is the number of families required to distinguish between the various models based on the mean number of affected children? We have investigated this question; the results are given in table 3. It is clear from this table that while the sample sizes (i.e. numbers of families to be sampled) for distinguishing between one-locus and two-locus models are within feasible limits at different levels of prevalence, this is not so for discriminating among multi-locus models based on mean number of affected offspring.

11. An empirical illustration of the likelihood method

Prelingual deafness refers to non-acquired hearing loss prior to attaining 3 years of age. Aetiological agents for acquired hearing loss include prenatal rubella infection, maternal drug therapy (e.g. quinine, thalidomide, etc.) during pregnancy, perinatal trauma, postnatal meningitis, middle-ear disease, etc. Deafness is also associated with many genetic syndromes, e.g. Waardenburg, Usher, etc. (Konigsmark and Gorlin 1976). There seems to be an extensive heterogeneity in the pattern of inheritance of hereditary, nonsyndromic deafness. While the pattern of inheritance in consanguineous families supports a one-locus autosomal recessive pattern of inheritance

Table 3. Numbers of N × N families, each ascertained through an affected offspring, required to distinguish between various models based on mean number of affected offspring per family.

	Prevalence								
	0·01			0·001			0·0001		
	Sibship size			Sibship size			Sibship size		
Models to be distinguished	2	4	6	2	4	6	2	4	6
One-locus vs two-locus	92	32	19	52	18	11	39	13	8
Two-locus vs three-locus	824	299	194	434	155	99	293	103	65
Three-locus vs four-locus	3518	1301	859	1922	701	457	1341	477	305
Four-locus vs five-locus	10,288	3844	2555	5871	2162	1420	4301	1544	992

with no sporadic cases (Stevenson and Cheeseman 1956; Chung *et al.* 1959; Chung and Brown 1970; Nance 1980), the inferred mode of inheritance in non-consanguineous families seems to be equivocal. In non-consanguineous families with positive family history, while Chung and Brown (1970) inferred the involvement of autosomal dominant genes with 80% penetrance and 12·8% sporadics, Nance (1980) inferred the involvement of autosomal recessive genes with about 20% sporadics. In non-consanguineous families with negative family history, while Chung *et al.* (1959), Chung and Brown (1970) and Nance (1980) have all inferred a single-locus autosomal recessive mode of inheritance, their estimates of the proportion of sporadic cases varied from 26% to 65%. The estimated number of recessive genes involved in deafness varies greatly. Using the theory of detrimental equivalents, Chung *et al.* (1959) estimated 36 genes, Chung and Brown (1970) estimated 5 genes, while Sank (1963) estimated a range of 45–6800 genes. It is also unclear whether these genes are all at one locus or at many loci. Chung *et al.* (1959) have stated that "it seems much more likely that there are many loci".

In view of these disconcerting discrepancies, we (Majumder *et al.* 1989) had undertaken a study on the genetics of prelingual deafness. Families (nuclear and extended) were ascertained through an affected offspring (the proband). Probands were drawn from various schools for the deaf in and around the city of Madras. Care was taken to ensure that all deafs were nonsyndromic and there was no acquired cause of deafness. Our data set comprised a total of 133 nuclear families, of which 83 had no parental consanguinity. In the remaining 50 consanguineous families, the parents were an uncle–niece pair in 19 families and a pair of first cousins in 31 families. In the 25 extended pedigrees that were collected, consanguinity was observed in 12 pedigrees; the remaining 13 pedigrees were non-consanguineous. Both parents of every deaf individual were of normal hearing in all nuclear families and also in the pedigrees.

Segregation analysis of these family data was performed using the likelihood method outlined earlier. The estimated prevalence of prelingual deafness used in this analysis was 58 per 100,000 births (i.e. approximately 0·0006). This estimate was obtained from the 1980 Census of the Physically Handicapped, and refers to the state of Tamil Nadu.

The results of segregation analysis are presented in figures 4 and 5 separately for nuclear families and for pedigrees. Likelihood computations for nuclear families—consanguineous and non-consanguineous—were performed using the methods described above. Likelihood computations for pedigrees were performed using the computer program package PAP (Hasstedt and Cartwright 1981). As is seen from these figures, for both nuclear families and for pedigrees, the non-genetic (sporadic) model (obtained by setting the gene frequency at each locus to 1, implying that all individuals are of the same genotype and thus there is no genetic predisposition) is clearly rejected. The one-locus dominant model, with or without sporadics, is also rejected for all families. For the one-locus recessive model, while increasing the proportion of sporadic cases improves the value of the likelihood function for nuclear families, for pedigrees the value of the likelihood function decreases with increase in the sporadics proportions. However, even for nuclear families, the value of the likelihood function under the one-locus recessive model even with 20% sporadic cases is smaller than that under the two-locus recessive model. The difference in the likelihood values between the two-locus recessive model and the three-locus recessive model

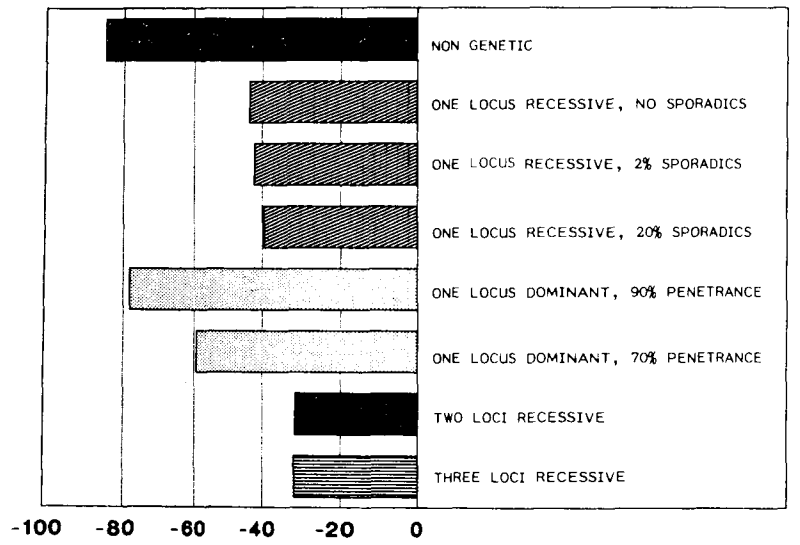

Figure 4. Log_{10}-likelihood values under different genetic models for 133 hearing × hearing nuclear families of prelingual deafness.

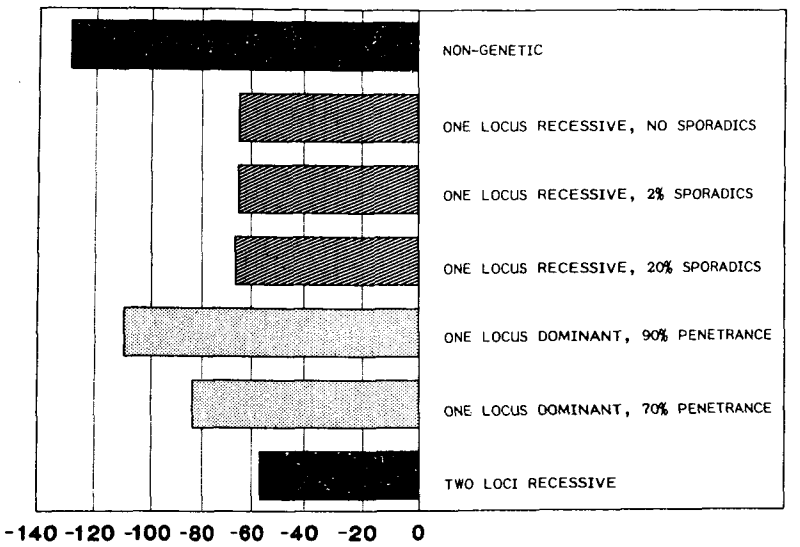

Figure 5. Log_{10}-likelihood values under different genetic models for 25 pedigrees of prelingual deafness.

is negligible for nuclear families. (Obtaining the value of the likelihood function for pedigrees under the three-locus model was computationally infeasible.) We thus conclude that the two-locus recessive model is the most parsimonious model for prelingual deafness.

12. Further complexity due to variable age of onset

When the disorder is not expressed at birth, and the age of onset is variable, further complexity arises because individuals may be of the susceptible genotype but may not have expressed the disorder. Unless this variability in age of onset is taken into account, the segregation probability will be underestimated. We present an approximate likelihood method for dealing with variable age of onset of N × N families ascertained through an affected offspring. Let:

$$\theta = \text{average segregation probability}$$
$$= \text{Prob (offspring is aabbcc... given that both parents are normal)}$$
$$= \sum_i \sum_j M_{ij}\theta_{ij}, \tag{18}$$

where M_{ij} and θ_{ij} are given by equations 7 and 8 respectively;

π = ascertainment probability;

n_i = total number of offspring of age i;

r_i = observed number of affected offspring of age i;

$n = \sum_i n_i$ = total number of offspring in the family;

$r = \sum_i r_i$ = total observed number of affected offspring in the family;

p_i = Prob (an individual of age i is affected given that he or she is aabbcc...).

The probabilities p_i are assumed to be known from age-specific and gender-specific estimates of prevalence of the disorder in the population. In practice, one may need to form age groups. Therefore, Prob (an offspring of age i is affected) $= p_i\theta$. Hence, the probability that a family with n_i offspring of age i will have r_i offspring of age i will be

$$\alpha_r \cdot P(n_i, r_i) = N \text{ (say)}, \tag{19}$$

where α_r is given by equation 9, and

$$P(n_i, r_i) = \prod_i \binom{n_i}{r_i} (p_i\theta)^r (1 - p_i\theta)^{n-r}. \tag{20}$$

Now, the probability that a family with n_i offspring of age i will have at least one affected offspring and is ascertained is

$$\sum_{r=1}^{n} \alpha_r \cdot \sum_{\substack{l_1,\ldots,l_g \\ l \le n, \Sigma l = r}} P(n_i, l_i) = D \text{ (say)}. \tag{21}$$

Hence, the conditional distribution of affected offspring in this family when it has been ascertained through an affected offspring (which is the likelihood of this family) is

$$L(\theta, \pi) = N/D. \tag{22}$$

If M families are sampled, the joint likelihood of all families is

$$L(\theta, \pi) = \prod_{m=1}^{M} L_m(\theta, \pi), \tag{23}$$

where $L_m(\theta, \pi)$ denotes the likelihood of the mth sampled family ($m = 1, 2, 3, \ldots, M$). Since there are two parameters to be estimated, the joint likelihood $L(\theta, \pi)$ may be numerically maximized to yield maximum likelihood estimates of θ and π.

Under single selection (i.e. $\pi \approx 0$), the likelihood simplifies to

$$L_m(\theta, \pi) = L_m(\theta) = \frac{r \cdot P(n_i, r_i)}{\sum_{r=1}^{n} r \sum_{l_1, \ldots, l_u} P(n_i, l_i)} \qquad (24)$$

which is independent of π.

13. Another empirical illustration

Vitiligo is an idiopathic hypomelanosistic dermatological disorder that is characterized by pale, milk-white maculae that tend to become progressive over time (Mosher *et al.* 1979). Positive family history, familial aggregation and concordance in monozygotic twins of vitiligo have been noted for a long time (Cockayne 1933; Merelender and Rywlin 1940; Mohr 1951; Siemens 1953; Behl 1955; Levai 1958; Lerner 1959; Mehta *et al.* 1973; Goudie *et al.* 1980; Hafez *et al.* 1983; Das *et al.* 1985a, b). However, systematic family studies to determine the mode of inheritance of vitiligo have sadly been lacking. To fill this gap we undertook an epidemiological and family study on vitiligo in Calcutta (Das *et al.* 1985a, b; Majumder *et al.* 1988). Based on these studies, which included clinical examination of 16,000 individuals and members of 298 families, the prevalence of vitiligo was estimated to be about 5 per 1000 individuals. The cumulative probability of onset of vitiligo by age of an individual was also estimated; these data are depicted in figure 6. The family data set included 188 N × N families each ascertained through an offspring with vitiligo. Using the p_i values depicted in figure 6, the joint likelihood of all families were calculated using equations 23 and 24 for various values of θ, assuming single selection. (We have actually investigated the validity of the

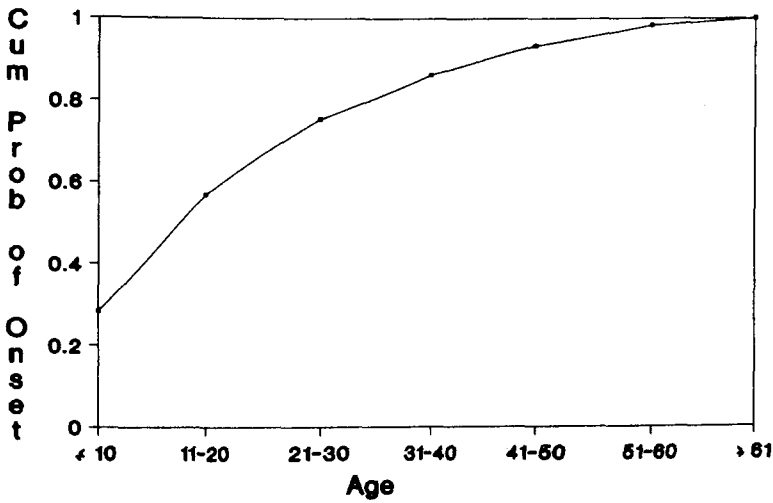

Figure 6. Cumulative probability of onset of vitiligo by age (in years).

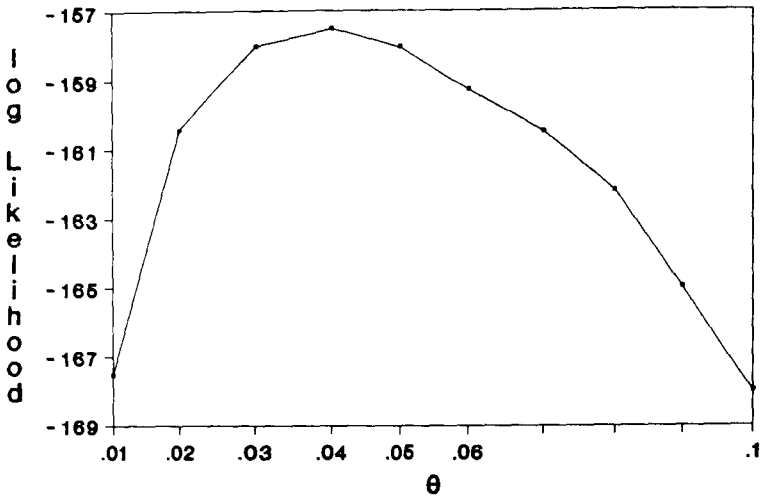

Figure 7. Log_e-likelihood values of 188 normal × normal families of vitiligo for different values of the segregation probability θ.

Table 4. Segregation probabilities (θ) for various values of k.

k	θ
1	0·2500
2	0·0993
3	0·0552
4	0·0368
5	0·0274

assumption of single selection; details are provided in Majumder *et al.* 1988.) The $\log_e$-likelihood values are plotted against θ in figure 7. It is seen that the likelihood at first monotonically increases with θ and then decreases. The maximum likelihood estimate of θ is $0{\cdot}039 \pm 0{\cdot}007$. What does this estimate of θ translate to in terms of the number of loci? In table 4 are given the θ, calculated using equation 18, for various values of k (the number of loci). In these calculations, we have used the estimated prevalence of vitiligo (Das *et al.* 1985a) of 0·00459, and have assumed equal gene frequencies at the various loci. Comparing the estimated value of θ with those given in table 4, it is seen that the data are compatible with a four-locus multiple recessive homozygosis model.

References

Behl P. N. 1955 Leucoderma and its treatment with Ammi Majus. *J. Indian Med. Assoc.* 24: 615–618

Chung C. S. and Brown K. S. 1970 Family studies of early childhood deafness ascertained through Clarke School for the Deaf. *Am. J. Hum. Genet.* 22: 630–644

Chung C. S., Robison O. W. and Morton N. E. 1959 A note on deaf mutism. *Ann. Hum. Genet.* 23: 357–366

Cockayne E. A. 1933 *Inherited abnormalities of the skin and its appendages* (Oxford: Oxford University Press)

Das S. K., Majumder P. P., Chakraborty R., Majumdar T. K. and Haldar B. 1985a Studies on vitiligo. I. Epidemiological profile in Calcutta, India. *Genet. Epidemiol.* 2: 71–78

Das S. K., Majumder P. P., Majumdar T. K. and Haldar B. 1985b Studies on vitiligo. II. Familial aggregation and genetics. *Genet. Epidemiol.* 2: 255–262

Goudie R. B., Goudie D. R., Dick H. M. and Ferguson-Smith M. A. 1980 Unstable mutations in vitiligo, organ-specific autoimmune diseases, and multiple endocrine adenoma peptic ulcer syndrome. *Lancet* 2: 285–287

Hafez M., Sharaf L. and El-Nabi S. M. A. 1983 The genetics of vitiligo. *Acta Derm.-Venereol. (Stockh.)* 63: 249–251

Hasstedt S. J. and Cartwright, P. E. 1981 PAP: pedigree analysis package. Tech. Rep. 13, Department of Medical Biophysics and Computing, University of Utah, Salt Lake City

Konigsmark B. W. and Gorlin R. J. 1976 *Genetic and metabolic deafness* (Philadelphia: W. B. Saunders Co.)

Lerner A. B. 1959 Vitiligo. *J. Invest. Dermatol.* 32: 285–310

Levai M. 1958 A study of certain contributory factors in the development of vitiligo in south Indian patients. *Arch. Dermatol.* 78: 364–371

Li C. C. 1953 Some general properties of recessive inheritance. *Am. J. Hum. Genet.* 5: 269–279

Li C. C. 1987 A genetical model for emergenesis: In memory of Laurence H. Snyder, 1901–86. *Am. J. Hum. Genet.* 41: 517–523

Li C. C. and Sacks L. 1954 The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. *Biometrics* 10: 347–360

Lykken D. T. 1982 Research with twins: the concept of emergenesis. *Psychophysiology* 19: 361–373

Majumder P. P., Das S. K. and Li C. C. 1988 A genetical model for vitiligo. *Am. J. Hum. Genet.* 43: 119–125

Majumder P. P., Ramesh A. and Chinnappan D. 1989 On the genetics of prelingual deafness. *Am. J. Hum. Genet.* 44: 86–99

Mehta N. R., Shah K. C., Theodore C., Vyas V. P. and Patel A. B. 1973 Epidemiological study of vitiligo in Surat area. *Indian J. Med. Res.* 61: 145–154

Merelander J. and Rywlin J. 1940 Heredity of acquired vitiligo (vitiligo in three generations). *Acta Derm.-Venereol. (Stockh.)* 21: 583–585

Mohr J. 1951 Vitiligo in a pair of monovular twins. *Acta Genet. Stat. Med.* 2: 252–255

Mosher D. B., Fitzpatrick T. B. and Ortonne J. P. 1979 Abnormalities of pigmentation. In *Dermatology in general medicine* (eds) T. B. Fitzpatrick, A. Z. Eisen, K. Wolf, I. K. Freedberg and K. F. Austen (New York: McGraw Hill Co.) pp. 582–590

Nance W. E. 1980 The genetic analysis of profound prelingual deafness. *Birth Defects, Orig. Artic. Ser.* 16(4): 263–269

Sank D. 1963 Genetic aspects of early total deafness. In *Family and mental health problems in a deaf population* (eds) J. D. Rainer, K. Z. Altshuler and F. J. Kallman (New York: New York Psychiatric Institute)

Siemens H. W. 1953 Die Zwillingspathologie der Vitiligo. *Acta Genet. Med. Gemellol. (Roma)* 2: 118–124

Stevenson A. C. and Cheeseman E. A. 1956 Hereditary deaf mutism with particular reference to Northern Ireland. *Ann. Hum. Genet.* 20: 177–231