American Journal of Medical Genetics 24:231–246 (1986)

A Bayesian Method for the Estimation of Penetrance: Application to Mandibulofacial and Frontonasal Dysostoses

André Rogatko, Carlos A.B. Pereira, and Oswaldo Frota-Pessoa

Institudo de Matemática e Estatística da Universidade de São Paulo (A.R., C.A.B.P.), Instituto de Biociências da Universidade de São Paulo (O.F.-P.), and CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico (A.R., C.A.B.P., O.F.-P.), Sâo Paulo, Brasil

We describe a Bayesian method of estimating penetrance from genealogical data. It consists of calculating the likelihood of the data alone to make inferences about penetrance without sample space considerations. The method is applied to mandibulofacial dysostosis giving a penetrance of 0.908 with 0.95 credible interval of [0.809; 0.972] and to frontonasal dysostosis giving a penetrance of 0.670 with 0.95 credible interval of [0.457; 0.851].

Key words: Bayes estimator, credible interval, frontonasal dysostosis, graph theory, likelihood, mandibulofacial dysostosis, penetrance

INTRODUCTION

The concept of penetrance was originally introduced by Vogt [1926] and reviewed and extended by Rogatko [1983] using transitional matrices to model the different types of penetrance. The penetrance concept, which has had wide application in medical genetics [Tanaka, 1967; Stevenson and Davison, 1970; Frota-Pessoa et al, 1976; Gollop, 1981; Opitz, 1981], will be of even greater value if incomplete penetrance is a widespread phenomenon. In fact, the complex nature of both gene and gene/environment interactions in humans suggests that incomplete penetrance is more frequent than is generally thought. Accordingly, more rigorous methods for determining penetrance must be developed so that, given any normal or pathological traits, the model best describing its inheritance can be determined. The evaluation of penetrance is also important in genetic counseling; it increases the precision of the recurrence risk estimates, although it does not clarify the detailed inheritance pattern.

Received for publication August 12, 1985; revision received September 25, 1985.

Address reprint requests to Dr. O. Frota-Pessoa, Instituto de Biociências-USP, Depto. de Biologia, CP.11461, 05421 São Paulo, Brasil.

TABLE I. Transitional Matrix [Rogatko, 1983]

Genotypes	Phenotypes		
	f_1	f_2	f_3
A_1A_1	1	0	0
A_1A_2	0	k	1−k
A_2A_2	0	0	1

Penetrance can be estimated either by simply determining the cases of nonpenetrance [Stern, 1973] or by population analysis, which requires more elaborate methods [Trankell, 1955; Elandt-Johnson, 1970; Suarez et al, 1976, 1977]. Both methods are based on the distribution of affected individuals in pedigrees. Several approaches have been developed for defects occurring either unilaterally or bilaterally in the body, such as retinoblastoma [Verschuer, 1937; Allen, 1952; Kealin, 1955; Knudson, 1971; Matsunaga, 1976, 1978, 1980] and the case of twins [Schinz, 1945; Lasker, 1947].

However, with the exception of the method of Stern [1973], none of these applies to genetic diseases, for which only a few cases have been described. Stern's method uses only information derived from proven nonpenetrant heterozygotes and the vertical connections in the pedigree. As such, it cannot determine the exact degree of precision of the estimate.

In this paper, we try to develop a method that estimates the penetrance of a given gene and establish, at the same time, the degree of uncertainty of the estimate taking into account all the information contained in the genealogies. The theoretical basis of this method warrants its application to any sample size. It is clear that, the smaller the sample considered, the less will be the amount of information contained in the data and, likewise, the smaller will be the reduction of the initial uncertainty.

METHOD

Model

We start from the transitional matrix shown in Table I (model VIII in Rogatko [1983]). In this matrix, k is the penetrance coefficient, the individuals exhibiting phenotype f_1 die, those having phenotype f_2 have a given Mendelian trait, and those with phenotype f_3 are normal for the trait considered. The fraction k of the heterozygotes have the trait and the fraction $1 - k$ and A_2A_2 homozygotes are both normal. Therefore, class f_3 is composed of phenotypically indistinguishable normal individuals of the types A_1A_2 and A_2A_2.

Genes following this pattern are often classified as dominant with incomplete penetrance. However, we prefer to call them *incomplete dominants* (or *codominants*), because, from a phenotypical point of view, three distinct classes occur.

Considering the adaptive values w_1, w_2, and w_3, corresponding to the phenotypic classes f_1, f_2, and f_3, we assume that $w_1 = 0$ and $w_2 = w_3$ (A_1A_1 is lethal, and the penetrant heterozygotes have the same adaptive value as the A_2A_2 homozygotes). The present method applies only to genes of small frequency. In such cases, the affected and their consanguineous relatives can be assumed to always cross with A_2A_2 individuals unless consanguinity is envolved in the cross.

In practice, the gene is recognized as being rare when, in a preliminary evaluation, the frequency of the affected phenotype is considered low in the population and high within the families of the propositi. Therefore, conditions that are rare because of a low penetrance are not considered here.

Likelihood

Several authors of different convictions have shown that statistical inference can be obtained from the analysis of likelihood functions. In the classical methods [Fisher, 1956; Birnbaum, 1962; Barnard, 1967; Edwards, 1972], total knowledge of the sample space (set of all possible results) is necessary. This stems from the fact that such methods are based on the probability distribution of data for each parameter value. In the case of the estimation of penetrance, it is difficult to define the sample space, because we cannot determine all possible family compositions resulting from the number and type of individuals in each generation of sibs, the number of generations, the types of crosses between individuals, and the presence or lack of consanguinity. Even if the sample space was determined, classical statistical theory would allow only imprecise inferences when the number of known genealogies of a given defect is small, because it relies on asymptotic results, which are reliable only in large samples.

These concerns led to the Bayesian approach for data analysis [Basu, 1975; Lindley and Phillips, 1976]. Contrary to the classical mode of inference, the Bayesian method takes into account only the results actually observed; it does not consider all the infinite possible observations that could have but did not occur. Besides, its application is not restricted by sample size. The Bayesian inference is composed of three entities: an *a priori* distribution, which indicates the initial amount of information held by the researcher; a likelihood function, which codes all relevant information contained in the data about the parameter in question; and an *a posteriori* distribution, which indicates the state of information about the parameter after the data have been analyzed.

When the researcher is faced with the first pedigree of a given disease, the *a priori* distribution of penetrance is uniform; no previous information exists. The *a posteriori* distribution evaluated for the first pedigree is the *a priori* distribution for the next one. In this iterative manner, information is accumulated and the final *a posteriori* distribution for the set of data available is reached. In practice this corresponds to constructing a single likelihood function using all pedigrees, associating with it a uniform *a priori* distribution, and computing the *a posteriori* distribution.

The method described in this paper is based on the analysis of the likelihood function constructed from all available pedigrees. This procedure corresponds to a Bayesian analysis with a uniform initial *a priori* distribution. Although using the maximum likelihood point estimate for penetrance, its precision is evaluated from the conditional (*a posteriori*) distribution of penetrance given the data. To the contrary, the classical methods evaluate the precision from the distribution of data given the parameter. This is the main practical difference between the two approaches, which, however, can lead to substantially different results.

Analysing the Genealogies

We call a *genealogy-generator* (GG): 1) an affected individual with no known affected ancestrals or 2) a normal couple with no affected ancestrals who are the

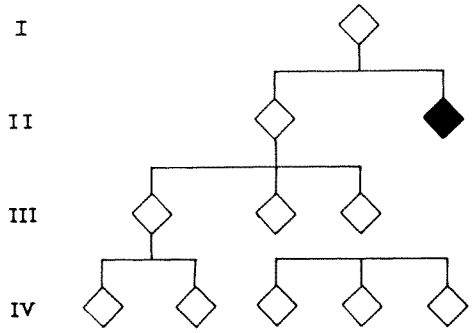

Fig. 1. Example of a tree of normal individuals (II-1–IV-5).

origin of two or more independent offspring lineages each of them containing affected individuals. In case 2, one of the members of the couple, with the exclusion of the other, is certainly a heterozygote.

It is convenient to represent in the pedigrees both members of a couple when, and only when, the couple is a GG or is an ancestor of a GG. All pertinent information in the genealogies will be condensed in a likelihood function L(k), which is a product of different factors.

The Evaluation of L(k)

When there is no information about the ancestrals of a GG, the latter contributes to the likelihood with a factor: 1) k, if it is affected, or 2) $1 - k$, if it is not affected. In the case when information about the ancestrals is available, those factors must be adjusted according to such information.

A GG descendent contributes to the likelihood with a factor: 3) k/2, if it is affected, 4) $(1 - k)/2$, if it is a nonpenetrant heterozygote (identifiable for having some affected descendent), or 5) $(2 - k)/2$, if it is a normal individual without offspring. If a normal GG descendent has only normal offspring its contribution must be adjusted accordingly.

The Adjustment of Likelihood Factors

The factor adjustments in the cases of information being available on GG ancestors or of a normal GG descendent having only normal offspring can be performed with reference to the following considerations. Consider a tree of normal individuals descending from a heterozygote known to be so because of having affected offspring (as in Fig. 1) or because of being affected. The likelihood factor corresponding to such a tree is calculated according to an order that works downwards and is called *node, left branch, right branch* in graph theory [Elson, 1975; Lewis and Smith, 1976]. The procedure is illustrated in Table II. It shows the possible homozygous or heterozygous condition of each individual depending on the condition ascribed to the ancestrals.

To transform the diagram shown in Table II into the algebraic expression of likelihood, operators $\rightarrow$ and *and* are replaced by multiplication and operator *or* by

TABLE II. Diagram of All Possible Alternatives for the Tree in Figure 1*

II-1 =	$h_1 \rightarrow$ or $H'_1 \rightarrow$	III-1-3 =	$h_0 \rightarrow$	IV-1-5 =	h_0
		III-1 =	$h_1 \rightarrow$ or $H'_1 \rightarrow$	IV-1-2 =	h_0
				IV-1 =	h_1 or H_1
				and IV-2 =	h_1 or H'_1
		and III-2 =	h_1 or H'_1		
		and III-3 =	$h_1 \rightarrow$ or $H'_1 \rightarrow$	IV-3-5 =	h_0
				IV-3 =	h_1 or H'_1
				and IV-4 =	h_1 or H'_1
				and IV-5 =	h_1 or H'_1

*The arrows indicate new generations, and the brackets gather sets of sibs or alternative possibilities of a same individual. h_0, homozygote derived from $A_2A_2 \times A_2A_2$; h_1, homozygote derived from $A_1A_2 \times A_2A_2$; H'_1, nonpenetrant heterozygote derived from $A_1A_2 \times A_2A_2$.

addition. Operators $\rightarrow$ and *and* have the same precedence and both precede operator *or*. This precedence can be changed by brackets.

According to these rules, the likelihood in the tree of normal individuals in Figure 1 is given by

$$\frac{1}{2} + \frac{1-k}{2}\left[\frac{1}{2} + \frac{1-k}{2}\left(\frac{2-k}{2}\right)^2\right]\left(\frac{2-k}{2}\right)\left[\frac{1}{2} + \frac{1-k}{2}\left(\frac{2-k}{2}\right)^3\right].$$

If a tree is formed by n normal individuals ($n > 1$), when each individual gives rise to only one descendent, the likelihod is

$$\frac{1}{2}\left[1 + \sum_{i=1}^{n-1}\left(\frac{1-k}{2}\right)^i\right] + \left(\frac{1-k}{2}\right)^n.$$

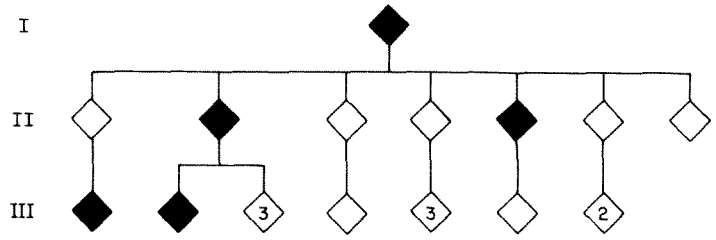

Fig. 2. Genealogy generated with k = 0.75.

The likelihood for any genealogy is given by

$$L(k) = L_{GG}(k/2)^{n_1}[(1 - k)/2]^{n_2}[(2 - k)/2]^{n_3} L_T,$$

where L_{GG} = GG factor (adjusted or not), n_1 = number of affected individuals, n_2 = number of heterozygotes, n_3 = number of normal individuals (with a heterozygote parent) with no offspring, and L_T = adjustment factor of trees of normal individuals, since the constants multiplying the likelihood function are merely a scale factor:

$$L(k) = L_{GG}k^{n_1}(1 - k)^{n_2}(2 - k)^{n_3}L_T.$$

An Example of Simulation

The genealogy shown in Figure 2 was produced from random number sequences, where the probability of an $A_1A_2 \times A_2A_2$ cross originating an A_1A_2 individual is 0.5, and the penetrance of the A_1A_2 heterozygote is 0.75. In this small-scale genealogy, four of the eight A_1A_2 individuals generated from I-1 manifested the gene. The likelihood of this genealogy is

$$L(k) = kk^4(1 - k)(2 - k)^5 L_T,$$

where

$$L_T = \left[\frac{1}{2} + \frac{1 - k}{2}\left(\frac{2 - k}{2}\right)\right]\left[\frac{1}{2} + \frac{1 - k}{2}\left(\frac{2 - k}{2}\right)^3\right]\left[\frac{1}{2} + \frac{1 - k}{2}\left(\frac{2 - k}{2}\right)^2\right]$$

$$L(k) = k^5(1 - k)(2 - k)^5\,[2 + (1 - k)\,(2 - k)] \times [4 + (1 - k)\,(2 - k)^2]\,[8 + (1 - k)\,(2 - k)^3];$$

this is graphically represented in Figure 3, giving a unimodal curve. By solving the equation $dL(k)/dk = 0$ through the Newton-Raphson iterative method (with the tolerance of convergence $\epsilon = 10^{-10}$), we obtain 0.618 for the mode of this function (Fig. 4). This value represents the penetrance that most probably would generate the genealogy in Figure 2.

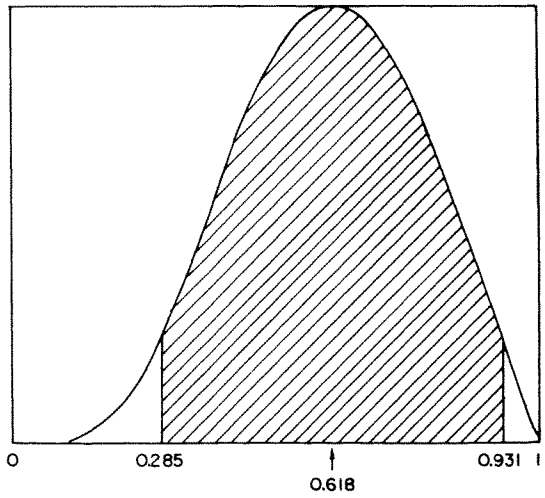

Fig. 3. Mode and 95% credible interval of the likelihood function of the simulated genealogy in Figure 2.

Credible Set

We shall now introduce the concept of "credible set" or interval in order to evaluate the degree of uncertainty of this parameter. The shaded area in Figure 3 represents a fraction $\alpha = 0.95$ of the total area beneath the curve:

$$\int_{a_1}^{a_2} L(k)dk = \alpha \int_0^1 L(k)dk.$$

In the particular case of Figure 3, $a_1 = 0.285$ and $a_2 = 0.931$. Furthermore, any ordinate within the shaded area is greater than any ordinate outside; that is, $|a_1 - a_2|$ is minimum, and $L(a_1) = L(a_2)$. The shaded area is called the *highest-density region, credible region,* or *Bayesian confidence interval* [Schmitt, 1969].

To include multimodal instances, the definition must be generalized. Given a family of sets

$$\mathscr{C} = \left\{C: \int_C L(k)dk = \alpha \int_0^1 L(k)dk\right\},$$

a credible set of α level is defined as the set of $C_0 \in \mathscr{C}$, so that

$$\int_C dk > \int_{C_0} dk \quad \forall\, C \in \mathscr{C}.$$

In spite of the analogy between the Bayesian credible interval and the confidence interval of classical statistics, they are quite distinct from a conceptual point of view.

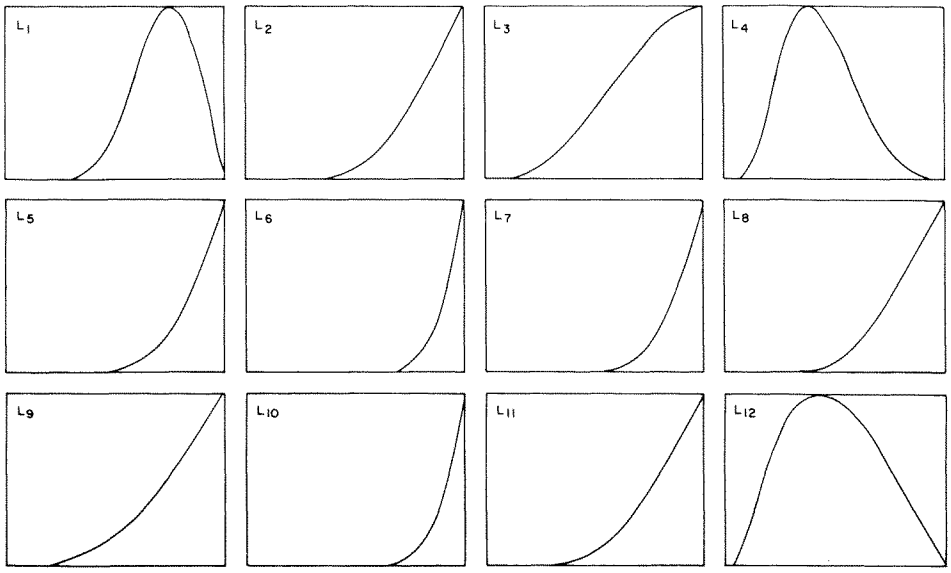

Fig. 4. Graphs of the likelihood functions L_i (ordinate) × penetrance (abscissa) for MFD.

A credible interval α means that the actual value of the parameter has a probability α of belonging to the set C_0. On the other hand, a confidence interval α means that, if the construction method was applied to all possible samples, then $100\alpha\%$ of the confidence intervals thus calculated should contain the actual value of the parameter. It should be noted that, for the observed sample, the confidence interval may or may not contain the actual value of the parameter.

Figure 3 represents the 95% credible region (shaded area) of the likelihood function for the simulated genealogy with k = 0.75 shown in Figure 2. The 95% credible interval is [0.285; 0.931] ($\epsilon = 10^{-10}$). This means that, on the basis of the information available for this genealogy, the actual penetrance value can be established to be between 0.285 and 0.931 with 95% probability. Before analyzing the data, our uncertainty was characterized by a uniform distribution in the [0,1] interval. Thus any credible interval corresponding to a 0.95 probability area is, a priori, an interval measuring 0.95 in the abscissae. With the information available in the sample, the initial uncertainty was reduced to [0.285; 0.931], and this interval is 0.646 long in the abscissae. The numeric method developed for the calculation of the credible interval is based on the half-interval search; the k values on the same ordinate are evaluated through the Newton-Raphson method, and the numeric integration of the likelihood function is obtained through Roemberg's quadrature [Kuo, 1972].

APPLICATION TO MANDIBULOFACIAL DYSOSTOSIS

In this section, the method described above will be applied to the study of penetrance in mandibulofacial dysostosis (MFD). MFD is a malformation syndrome characterized by incomplete or abnormal embryonic formation of the several facial structures in past derived from first and second branchial arches. The cause is considered to be an autosomal dominant gene with incomplete penetrance. Clinical

TABLE III. Likelihoods (L_i) of Genealogies With MFD in the Literature

Debusmann [1940]	$L_1 = k^{10}(1 - k)(2 - k)^{12}$
Leopold et al [1945]	$L_2 = k^6(2 - k)^4$
Brohm and Kluska [1947]	$L_3 = k^3(2 - k)^3$
Waardenburg and Navis [1949]	$L_4 = k^4(1 - k)(2 - k)^{14}$
Szlazak [1953]	$L_5 = k^8(2 - k)^4$
Rovin et al [1964]	$L_6 = k^{19}(2 - k)^{12}[4 + (1 - k)(2 - k)^2]$
Fazen et al [1967]	$L_7 = k^{10}(2 - k)^5$
Farrar [1967]	$L_8 = k^8(2 - k)^6[8 + (1 - k)(2 - k)^3]^2[16 + (1 - k)(2 - k)^4]$
Kirkham [1970]	$L_9 = k^3(2 - k)$
Le Marec et al [1974]	$L_{10} = k^{10}(2 - k)^2[2 + (1 - k)(2 - k)]$
Partsch and Husle [1975]	$L_{11} = k^5(2 - k)^3$
Gollop [1981]	$L_{12} = k^2(1 - k)(2 - k)^2[4 + (1 - k)(2 - k)^2][8 + (1 - k)(2 - k)^3]$

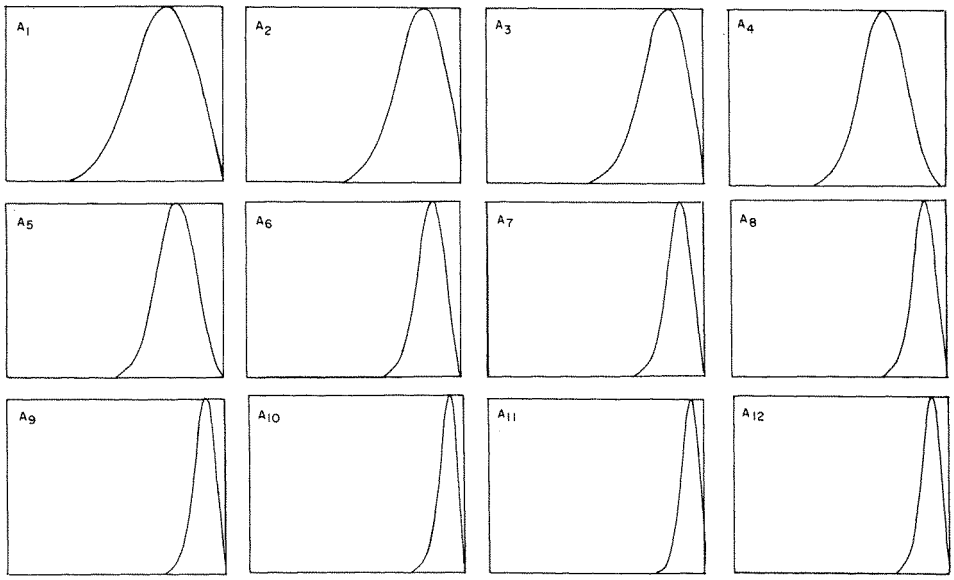

Fig. 5. Graphs of the accumulated likelihood functions A_i (ordinate) × penetrance (abscissa) for MFD.

and genetic information on this condition can be found in Gollop [1981], which is the source of the genealogies used here.

Table III shows the likelihoods of the 12 genealogies with MFD included in the review of Gollop [1981]. These genealogies are chronologically arranged, according to publication dates. The graphs in Figure 4 represent the likelihood functions L_i (ordinate) × penetrance (abscissa). Those in Figure 5 represent the accumulated likelihood functions

$$A_i = \prod_{j=1}^{i} L_j \text{ (ordinate)} \times \text{penetrance (abscissa)}.$$

The graphs in Figure 5 show that, the more information (genealogies) is added to the likelihood function, the more the curve peak narrows, thus delimiting an

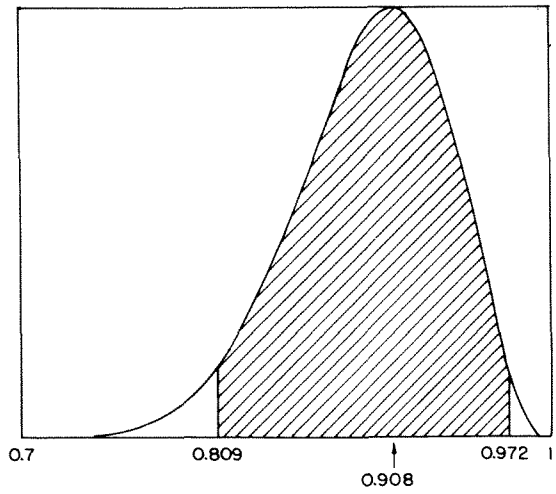

Fig. 6. Mode and 95% credible interval of MFD likelihood function A_{12}.

increasingly narrower credible interval and improving the precision of the penetrance estimate.

The likelihood function containing all information on penetrance of MFD is

$$\begin{aligned} L(k) &= A_{12} \\ &= k^{88}(1-k)^3(2-k)^{68}[2+(1-k)(2-k)] \\ &\times [4+(1-k)(2-k)^2]^2[8+(1-k)(2-k)^3]^3 \\ &\times [16+(1-k)(2-k)]^4. \end{aligned}$$

Its mode is 0.908, and its 95% credible interval is [0.809; 0.972] ($\epsilon = 10^{-10}$), which measures 0.163, as shown in Figure 6.

PENETRANCE AND INBREEDING

The possibility of two heterozygotes mating exists whenever consanguineous unions occur in genealogies of genetic diseases following the model described in the Method section under model. In that the homozygotes A_1A_1 were assumed to be inviable, the probability that a child of heterozygous parents will be a normal homozygote, an affected heterozygote, or a nonpenetrant heterozygote is 1/3, 2k/3 or 2(1 − k)/3, respectively.

Table IV shows likelihoods for hypothetical genealogies with consanguineous marriages. The likelihood, in these instances, was evaluated by listing all the possible combinations, calculating the likelihoods of each configuration, and summing.

In genealogies with several consanguineous marriages, the enumerations become tedious, and, besides, there are many possibilities of error. For this reason, we have developed an algorithm for the calculation of likelihood that scans and lists

TABLE IV. Likelihoods for Genealogies With Consanguineous Marriages*

Type of consanguineous marriage	Genealogy	Likelihoods (furthest descendant [III-1, IV-1, or V-1] is)		
		Affected	Nonpenetrant heterozygote	Normal (homozygote or heterozygote)
A. Siblings	II III	$k(2k^2-7k+5)/12$	$(-2k^3+9k^2-12k+5)/12$	$(-2k^3+10k^2-17k+12)/12$
B. Uncle × niece (or aunt × nephew)	II III IV	$k(-2k^3+9k^2-15k+8)/24$	$(2k^4-11k^3+24k^2-23k+8)/24$	$(k^4-6k^3+15k^2-19k+12)/12$
C. First cousins	II III IV	$k(2k^4-11k^3+27k^2-29k+11)/48$	$(-2k^5+13k^4-38k^3+56k^2-40k+11)/48$	$(-4k^5+25k^4-69k^3+106k^2-97k+51)/48$
D. First cousins once removed	II III IV V	$k(-2k^5+13k^4-38k^3+62k^2-52k+17)/96$	$(2k^6-15k^5+51k^4-100k^3+114k^2-69k+17)/96$	$(2k^6-16k^5+59k^4-131k^3+187k^2-173k+96)/96$
E. Second cousins	II III IV V	$k(k^6-9k^5+33k^4-68k^3+81k^2-51k+13)/96$	$(-k^7+10k^6-42k^5+101k^4-149k^3+132k^2-64k+13)/96$	$(-2k^7+18k^6-75k^5+196k^4-348k^3+426k^2-359k+192)/192$

*The GG is always heterozygote and his/her furthest descendant is affected, nonpenetrant heterozygote, or normal homozygote or heterozygote.

TABLE V. Method for the Calculation of Likelihood in Genealogies With Consanguineous Marriages*

$I\text{-}1 = A_1A_2 \rightarrow$ [II-1 = [$h_1 \rightarrow$ III-1 = h_0 I_1 ; or ; $H'_1 \rightarrow$ III-1 = [h_1 I_2 ; or ; H'_1 I_3]

Consanguinity list (for explanation, see text):
-III-2: III-1 = [h(1,2) or H'(3)]
$I_1 = 1/2$; $I_2 = (1 - k)/4$; $I_3 = (1 - k)^2/4$

and
II-2 = [$h_1 \rightarrow$ III-2 = h_0 and III-1 = [$h \rightarrow$ IV-1 = h_0 ; or ; $H' \rightarrow$ IV-1 = [h_1 ; or ; H_1] ; β_1]

or
$H'_1 \rightarrow$ III-2 = [h_1 and III-1 = [$h \rightarrow$ IV-1 = h_0 ; or ; $H' \rightarrow \beta_1$] ; or ; H'_1 and III-1 = [$h \rightarrow \beta_1$; or ; $H' \rightarrow$ IV-1 = [h_2 ; or ; H'_2]]

$\beta_1 = (2 - k)/2$:

Therefore, the likelihood is

$$L(k) = 1/2 \cdot 1 \cdot [I_1 + I_2) \cdot 1 + I_3 \cdot \beta_1] + (1 - k)/2 \cdot \{1/2 \cdot [(I_1 + I_2) \cdot 1 + I_3 \cdot \beta_1] + (1 - k)/2 \cdot [(I_1 + I_2) \cdot \beta_1 + I_3 \cdot (3 - 2 \cdot k)/3]\} = (-4k^5 + 25k^4 - 69k^3 + 106k^2 - 97k + 51)/48$$

*Marriage between first cousins when individual IV-1 is normal (see Table IV). h_2, homozygote derived from $A_1A_2 \times A_1A_2$; H'_2 nonpenetrant heterzygote derived from $A_1A_2 \times A_1A_2$ (see Table II).

genealogies in the way shown in Table II. For further explanation on the use of this algorithm, see Table V.

The rules used in this procedure are: 1) The genealogy is listed according to the node left branch, right branch order, using the abbreviations in Table I. 2) Concomitantly, a list of consanguinity is elaborated. This list is consulted for each new individual listed. Three possibilities arise, at this stage: 1) If the individual is not on the consanguinity list up to this point, and has not bred with a consanguineous relative the listing is continued. 2) If the individual is not in the consanguinity list but has bred with a consanguineous relative, then the identification of his/her partner, the identification, the genotype and the phenotype codes of the individual in question, and the interim likelihood index are included in the consanguinity list. By *interim likelihood* (I_i, $i = 1, 2 \ldots$), we mean the likelihood obtained by scanning the genealogy in an ascending order, through the operators $\rightarrow$ and *and*, starting from the individual in question up to the GG. If in the ascendant scanning alternatives linked by the *and* operator occur, and these alternatives have no connection with the other paths that form the consanguineous loop, they will take part wholly (ie, by scannings

TABLE VI. Likelihoods (L_i) of Genealogies With FND in the Literature

Boo-Chai [1965]	$L_1 = k^3(1 - k)(2 - k)^4$
Francesconi and Fortunato [1969]	$L_2 = k^3(1 - k)$
	$L_3 = k^5(1 - k)(2 - k)^4[8 + (1 - k)(2 - k)^3]$
Warkany et al [1973]	$L_4 = k^2(1 - k)(2 - k)$
Fox et al [1978]	$L_5 = k^2(1 - k)$

downward by the $\rightarrow$, *and*, and *or* operators) in the I_i. 3) If an individual is on the consanguinity list, his/her likelihood is multiplied by the sum of the interim likelihoods for each genotype and phenotype of the partner, and the scanning of the genealogy is continued, with the listing of all possible alternatives.

When calculations are done by hand, it is convenient to create auxiliary variables (β_i, i = 1, 2 . . .) that accumulate the likelihood values for repeating parts of the genealogy. These variables are represented in the listing by larger triangular arrows. Table V shows an example of this algorithm used in the situation already shown in Table IV (marriage between first cousins when individual IV-1 is normal).

APPLICATION TO FRONTONASAL DYSOSTOSIS (FND)

FND is a congenital face malformation resulting from disturbances in the embryogenesis of the frontonasal process. Its inheritance is considered as autosomal dominant with incomplete penetrance. Our estimate of its penetrance was based on the six genealogies reviewed in Gollop [1981].

Table VI shows the likelihood of the five genealogies without consanguinity. The graphs in Figure 7 represent the likelihoods functions L_1 (ordinate) $\times$ penetrance (abscissa). Those in Figure 8 represent the accumulated likelihood functions

$$A_i = \prod_{j=1}^{i} L_j \text{ (ordinate)} \times \text{penetrance (abscissa)}.$$

It can be seen that the graphs L_i are quite even in pattern as far as the location of the mode is concerned and that the peaks of the A_i curves narrow as more information is added to the likelihood function.

The likelihood function containing all information on the penetrance in FND in these five genealogies is

$$L(k) = A_5 = k^{15}\,(1 - k)^5\,(2 - k)^9\,[8 + (1 - k)\,(2 - k)^3].$$

The mode in this function (A_5 in Fig. 9) is 0.670, and the 95% credible interval is [0.457; 0.851], measuring 0.394.

The remaining FND genealogy, described by Moreno-Fuenmayor [1980], has many consanguineous marriages. Its likelihood is computed following the rules introduced in the Penetrance and Inbreeding section. Its final expression and the calculation of it are quite long and will be omitted here but are available to those interested. We present the graph representing the likelihood function of Moreno-

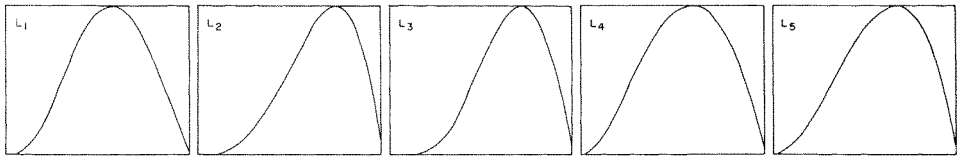

Fig. 7. Graphs of the likelihood functions L_i (ordinate) $\times$ penetrance (abscissa) for FND.

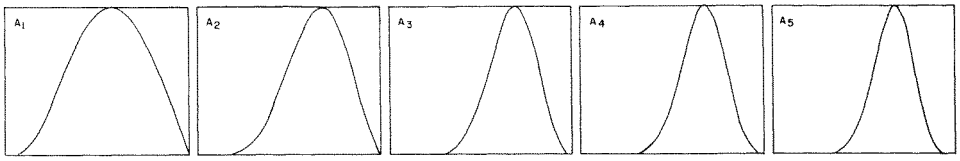

Fig. 8. Graphs of the accumulated likelihood functions A_i (ordinate) $\times$ penetrance (abscissa) for FND.

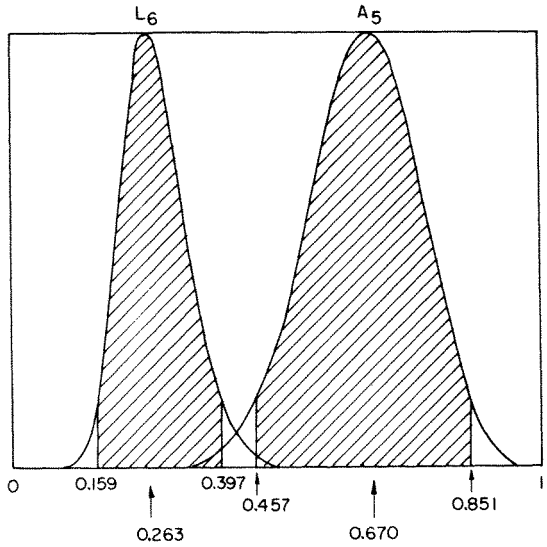

Fig. 9. Modes and 95% credible intervals of MFD likelihood functions L_6 and A_5.

Fuenmayor's pedigree (L_6 in Fig. 9), with mode 0.263 and 95% credible interval [0.159; 0.397] ($\epsilon = 10^{-10}$), measuring 0.238.

As can be seen in Figure 9, the L_6 and A_5 credible sets are disjoint. This result might be accounted for by a lack of information about generations I and II in Moreno-Fuenmayor's genealogy, in which no affected individuals are recorded, whereas a number of them appear in generations III–V. This would tend to make the calculated

penetrance an underestimate. On the other hand, we could be dealing with genetic heterogeneity, the Moreno-Fuenmayor pedigree representing a variant with lower penetrance.

REFERENCES

Allen G (1952): The meaning of concordance and discordance in estimation of penetrance and gene frequency. Am J Hum Genet 4:155–172.

Barnard GA (1967): The use of likelihood function in statistical practice. In: "Berkeley Symposium on Mathematical Statistics and Probability 5, 1965, Proceedings." Berkeley: University of California Press, pp 27–40.

Basu D (1975): Statistical information and likelihood. Sankhyã Series A 37(1):1–71.

Birnbaum A (1962): On the foundations of statistical inference. J Am Statist Assoc 57:269–326.

Boo-Chai K (1965): The bifid nose. With a report of 3 cases in siblings. Plast Reconstr Surg 36:626–628.

Brohm F, Kluska V (1947): Dysostosis mandibulofacialis (Franceschetti-Zwahlen). Lek Listy 2:1–16.

Cannings C, Thompson EA, Skolnick M (1980): Pedigree analysis of complex models. In Mielke JH, Crawford MH (eds): "Current Developments in Anthropological Genetics." New York: Plenum, Vol. 1, p 436.

Debusmann K (1940): Familiäre Kombinierte Gesichtsmissbildung im Bereich des ersten Viszeralbogens. Arch Kinderheilkd 120:133–139.

Edwards AWF (1972): "Likelihood: An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference." Cambridge: Cambridge University Press.

Elandt-Johnson RC (1970): Segregation analysis for complex models of inheritance. Am J Hum Genet 22:129–144.

Elson M (1975): "Data Structures." Chicago: Science Research Associates.

Farrar JE (1967): Mandibulo-facial dysostosis. Br J Ophthalmol 51:132–135.

Fazen LE, Elmore JBA, Nadler HL (1967): Mandibulo-facial dysostosis. Am J Dis Child 113:405–410.

Fisher RA (1956): "Statistical Methods and Scientific Inference." London: Oliver and Boyd.

Fox JW, Golden GT, Edgerton MT (1978): Frontonasal dysplasia with alar clefts in two sisters. Plast Reconstr Surg 43:481–495.

Francesconi G, Fortunato G (1969): Median disraphia of the face. Plast Reconstr Surg 43:481–491.

Frota-Pessoa O, Otto PA, Olivares Plaza JR (1976): The variations of recurrence risks according to penetrance for families with isolated cases of autosomal dominant conditions. J Hered 67(4):256.

Gollop TR (1981): Estudo genético-clínico das disostoses madibulofacial e frontofacionasal. São Paulo: IBIOC-USP (Tese-Doutoramento).

Kealin A (1955): Statistische Prüf-und Schätzverfahren für die relative Häufigkeit von Merkmals-trägern in Geschwisterreihen bei einem der Auslese unterworfenen Material mit Anwendung auf das Retino-blastom. Arch Julius-Klaus Stift 30:263–485.

Kirkham TH (1970): Mandibulofacial dysostosis with ectopia lentis. Am J Ophthalmol 70:947–949.

Knudson AG (1971): Mutation and cancer: Statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820–823.

Kuo SS (1972): "Computer Applications of Numerical Methods." Reading, PA: Addison Wesley.

Lasker GW (1947): Penetrance estimated by frequency of unilateral occurrences and by discordance in monozygotic twins. Hum Biol 19:217–230.

Le Marec B, Bourdinière J, Jezequel J, Senecal J (1974): L'antecipation et le syndrome de Franceschetti-Zwahlen-Klein. J Genet Hum 22:7–19.

Leopold IH, Mahoney JF, Price ML (1945): Symmetric defects in the lower lids associated with abnormalities of the zygomatic processes of the temporal bones. Arch Ophthalmol 34:210–218.

Lewis TG, Smith MZ (1976): "Applying Data Structures." Boston: Houghton Miffflin.

Lindley DV, Phillips LD (1976): Inference for a Bernoulli process (a Bayesian view). Am Statist 30(3):112–119.

Matsunaga E (1976): Hereditary retinoblastoma: Penetrance, expressivity and age of onset. Hum Genet 33:1–15.

Matsunaga E (1978): Hereditary retinoblastoma: Delayed mutation or host resistance. Am J Hum Genet 30:406–424.

Matsunaga E (1980): On estimating penetrance of the retinoblastoma gene. Hum Genet 56:127–128.
Moreno-Fuenmayor H (1980): The spectrum of frontonasal dysplasia in an inbred pedigree. Clin Genet 17:137–142.
Opitz JM (1981): Some comments on penetrance and related matters. Am J Med Genet 8:265–274.
Partsch CJ, Husle M (1975): Verschiedene Schwerhörigkeitsformen innerhalb einer Franceschetti Familie. Laryngol Rhinol Otol 54:385–388.
Rogatko A (1983): Diallelic model combining selection and incomplete penetrance. J Theor Biol 102:1–6.
Rovin S, Dachi SF, Barenstein DB, Cotter WB (1964): Mandibulofacial dysostosis, a familial study of five generations. J Pediatr 65:215–221.
Schinz HR (1945): Konkordanz, Diskordanz und Penetranz bei eineiigen Zwillingen (Versuch einer elementaren Darstellung). Arch Julius-Klaus Stift 20:2–25.
Schmitt SA (1969): "Measuring Uncertainty. An Elementary Introduction to Bayesian Statistics." California: Addison-Wesley.
Stern C (1973): "Principles of Human Genetics, ed 3." San Francisco: WH Freeman.
Stevenson AC, Davison BCC (1970): "Genetic Counselling." London: Heinemann.
Suarez BK, Reich T, Trost J (1976): Limits of the general two-allele single locus model with incomplete penetrance. Ann Hum Genet 40:231–243.
Suarez BK, Fishman PM, Reich T (1977): Estimating the parameters of the incompletely penetrant single locus model using multiple populations. Hum Hered 27:336–351.
Szlazak J (1953): Treacher-collins syndrome. Can Med Assoc J 69:274–279.
Tanaka K (1967): A comment on the estimation of genetic risks: Use of Bayes' theorem. Proc Jpn Acad 43(3):214–219.
Trankell A (1955): Penetrance calculus in population genetics. Am J Hum Genet 7:264–276.
Verschuer O (1937): "Erbpathologie." Dresden and Leipzig: Steinkopff.
Vogt O (1926): Psychiatrisch wichtige Tatsachen der zoologisch-botanischen Systematik. Z Ges Neurol Psychiatr 101:805–832.
Waardenburg P, Navis G (1949): Beitrag zur Kenntnis der Dysostosis mandibulofacialis und ihre Ontogenese. Acta Genet Statist Med 1:219–244.
Warkany J, Bofinger MK, Benton C (1973): Median facial cleft syndrome in half-sisters. Dilemmas in genetic counselling. Teratology 8:273–285.

Edited by John M. Opitz and James F. Reynolds