Am. J. Hum. Genet. 42:889–897, 1988

Segregation Analysis with Uncertain Ascertainment: Application to Fanconi Anemia

André Rogatko* and Arleen D. Auerbach†

*Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York; and †The Rockefeller University, New York, NY

Summary

A Bayesian solution for making inferences about segregation parameters with no information about the ascertainment is presented. Inferences about the segregation probability and the probability of being sporadic are made through the posterior marginal distribution of these parameters after integrating out the ascertainment probability, the nuisance parameter. The method was tested with real and simulated data and performed well. Original Fanconi anemia data, for which no information about the ascertainment was available, were then analyzed, with results that confirmed a monogenic autosomal recessive mode of inheritance.

Introduction

Human genetic data and ascertainment are tightly interconnected. Fisher (1934) was the first to define the probability of ascertainment, π, noticing that the segregation probability, p, is dependent on this ancillary parameter. Assuming a constant ascertainment probability (or a partition of the data set with different but constant ascertainment probabilities in each subset) and that the probands were independently ascertained, he proposed an estimator of π. These assumptions of constancy and independence of ascertainment have since been criticized as unrealistic (Stene 1977; Ewens and Shute 1986*a*, 1986*b*; Greenberg 1986). Nevertheless, the models adopting these assumptions generally produce consistent results. Haldane (1938) presented the classical model of segregation analysis with p and π, deriving the formula for the limiting cases of π ($\pi \rightarrow 0$, $\pi = 1$). Morton (1959) introduced the probability of being sporadic, x, into Haldane's model to estimate the frequency of sporadic cases, in which the affected status is due to mutation, phenocopies, polygenic complexes, etc.

The genetic parameters of main interest are p and x. The nuisance parameter is π, although its estimate may provide information about the prevalence of the disorder. The usual way of eliminating nuisance parameters, by large-sample theory, is to estimate the nuisance parameter away and substitute the unknown parameter π by its estimated value $\hat{\pi}$. Therefore, to have a successful analysis, it is necessary to gather information not only about the number of sibships of size s with r affected sibs (SR table) but also about π itself. Under the assumptions of constant and independent π, information about the ascertainment probability can be provided by the number of sibships with r affected sibs when a of them are probands (RA table) or, if there are independent sources of ascertainment (such as physicians, hospitals, birth and death certificates, and patient associations), by the number of t ascertainments that a proband has (T table).

Because of the inherent problems in human genetics of collecting data, sometimes the only information available is the SR table. Under these circumstances, the solution of conditioning the data on the limiting values of π and concluding that the most likely true values of p and x are bounded by these estimates is of difficult statistical interpretation. Another possibility

Received September 28, 1987; revision received December 10, 1987.

Address for correspondence and reprints: Dr. A. Rogatko, Box 44, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021.

0002-9297/88/4206-0013$02.00

is to maximize the joint distribution of p, x, and π, but this procedure is not supported theoretically because π is ancillary to this information (Stene 1981) and because, in practice, π generally converges to a boundary value.

To solve the problem of lack of information about the ascertainment, a Bayesian method for making inferences about the segregation parameters is presented. The performance of the method is tested with data with known π (cystic fibrosis and simulated data). In these cases, by assuming that π is unknown, the agreement of the recovered estimates with the true values can be examined. After the evaluation of the method, it is applied to Fanconi anemia (FA) data, for which no information about π is available. The investigation of whether FA might be defined by a cellular marker in homozygous cells was the primary motivation for the development of the method here presented. The question to be answered is whether classification of patients on the basis of this marker results in groups that may be considered as distinct entities on the basis of genetic segregation.

Material and Methods

Cystic Fibrosis

Cystic fibrosis data from the studies of Crow (1965), Danks et al. (1965), and Wright and Morton (1968) are used to test the performance of the method because, in addition to the SR tables, information about π is also available. In the Crow paper 71 affected individuals are reported and the RA table is presented. In the Danks et al. paper 213 affected individuals and both RA and T tables are available. In the Wright and Morton paper 21 affected individuals and both RA and T tables are reported. Data from the three studies are summarized in the Wright and Morton paper.

Simulated Data

Deterministic and pseudorandom simulated samples are also utilized to test the performance of the method. Segregation distributions of affected individuals were generated by formula (1) (see below), choosing particular values of s, p, x, and π. Deterministic samples are obtained by multiplying the generated probabilities by an arbitrary sample size N.

Pseudorandom simulated samples are obtained by using the segregation probabilities in a multinomial random deviate generator from the IMSL Library (1982) for a given N.

Fanconi Anemia

FA is an autosomal recessive disorder characterized by pancytopenia, congenital abnormalities, chromosome instability, and increased predisposition to cancer (Schroeder et al. 1964; Fanconi 1967; Alter and Potter 1983). The FA phenotype is highly variable, and the diagnosis of the syndrome on the basis of clinical manifestations alone is often difficult (Glanz and Fraser 1982). On the other hand, sensitivity to the clastogenic effect of DNA cross-linking agents such as diepoxybutane (DEB) is remarkable in homozygous FA cells (Auerbach and Wolman 1976; Auerbach et al. 1981). Therefore, clastogen-induced chromosomal breakage can be used as a cellular marker for the diagnosis of FA in presymptomatic cases or even prenatally (Auerbach et al. 1985, 1986).

The FA data analyzed came from two sources: the International Fanconi Anemia Registry at The Rockefeller University (Auerbach et al. 1988) and literature data from Schroeder et al. (1976). The International Fanconi Anemia Registry consists of possible FA cases reported mainly by physicians at medical centers specializing in the treatment of aplastic anemia. The probands and all siblings are tested for sensitivity of cultured lymphocytes to DEB and are classified as DEB^+ or DEB^-. On the basis of clinical evidence (Auerbach et al. 1988), we are classifying DEB^+ patients as affected with FA and DEB^- patients as nonaffected. In the literature data, FA is defined primarily on the basis of clinical symptoms.

The partition of affected individuals in the FA data is 88 in the DEB^+ group, 31 in the DEB^- group, and 86 in the LITERATURE group. The SR tables for the three groups are given in table 1.

Posterior Marginal Distributions

If one assumes a constant π and that the probands were independently ascertained, the probability of r affected individuals, given a sibship of size s and given p, x, and π (Morton 1959), is

$$Pr\{r|s,p,x,\pi\} = \begin{cases} \dfrac{sp\pi\,[x + (1-x)\,(1-p)^{s-1}]}{xsp\pi + (1-x)\,[1-(1-p\pi)^s]} & \text{for } r = 1 \\ \dfrac{(1-x)\,\binom{s}{r}\,p^r\,(1-p)^{s-r}\,[1-(1-\pi)^r]}{xsp\pi + (1-x)\,[1-(1-p\pi)^s]} & \text{for } r > 1 \end{cases} \quad (1)$$

Table 1

Number of *r* Affected among *s* Observed Sibs, for DEB$^+$, DEB$^-$, and LITERATURE

	r			
s AND GROUP	1	2	3	6
2:				
DEB$^+$	35	6		
DEB$^-$	12	5		
LITERATURE	22	4		
3:				
DEB$^+$	14	11		
DEB$^-$	9	1		
LITERATURE	12	4	2	
4:				
DEB$^+$	8	7	2	
DEB$^-$	2			
LITERATURE	5	4	1	
5:				
DEB$^+$		2		
DEB$^-$	1			
LITERATURE	5	6	3	
6:				
LITERATURE	5			
7:				
DEB$^+$			1	
LITERATURE	2	4	3	
8:				
DEB$^+$		2		
DEB$^-$				
LITERATURE				1
9:				
DEB$^-$			1	
10:				
LITERATURE	1			
12:				
LITERATURE			1	
14:				
LITERATURE				1

Let S be the maximal sibship size in the sample and a_{rs} be the sampled number of sibships of size s, each with r affected sibs. The likelihood of the sample, the triangular matrix a_{rs} (the SR table), given p, x, and π, is proportional to

$$Pr\{a_{rs}|p,x,\pi\} \propto \prod_{s=2}^{S} \prod_{r=1}^{s} Pr\{r|s,p,x,\pi\}^{a_{rs}}$$

In Bayesian theory, the unknown prameters to be estimated are associated with a probability distribution that represents our uncertainty about them. Technically speaking, unknown parameters receive the same treatment as random variables. Our aim is to make inferences on the parameters p and x. This can be achieved by analyzing either their joint distribution or the two univariate distributions of p and x alone. If one is to obtain the joint posterior marginal distribution (PMD) of p and x, π is the nuisance parameter to be eliminated. If one is to obtain the univariate PMD of p, the nuisance parameters are x and π. To get the univariate PMD of x, both p and π ought to be eliminated. The Bayesian method to eliminate nuisance parameters (Basu 1977) is as follows: fix a prior distribution, compute the posterior distribution, integrate out the nuisance parameter from the posterior distribution to obtain the PMD of the parameter of interest, and make inferences on the basis of the PMD. Let $Pr\{p,x,\pi\}$ be the prior distribution for p, x, and π. Therefore, the posterior distribution for p, x, and π is $Pr\{p,x,\pi|a_{rs}\} \circ Pr\{a_{rs}|p,x,\pi\} \cdot Pr\{p,x,\pi\}$. The joint PMD of p and x is

$$Pr\{p,x|a_{rs}\} = \int_0^1 Pr\{p,x,\pi|a_{rs}\}\, d\pi.$$

The univariate PMD of p is

$$Pr\{p|a_{rs}\} = \int_0^1 \int_0^1 Pr\{p,x,\pi|a_{rs}\}\, dx\, d\pi,$$

and the univariate PMD of x is

$$Pr\{x|a_{rs}\} = \int_0^1 \int_0^1 Pr\{p,x,\pi|a_{rs}\}\, dp\, d\pi.$$

To obtain the standardized PMD, the PMD is divided by the scale factor

$$\int_0^1 \int_0^1 \int_0^1 Pr\{p,x,\pi|a_{rs}\}\, dp\, dx\, d\pi.$$

All the integrations are performed numerically by an adaptive Romberg quadrature (Boor 1971).

When information about π is available (RA or T tables), informative priors may be used. If one assumes a constant π and that the probands were independently ascertained, the distribution of a probands among r affected in ascertained sibships, given π, is a truncated binomial (Fisher 1934):

$$Pr\{a|r,\pi\} = \frac{\binom{r}{a} \pi^a (1-\pi)^{r-a}}{1-(1-\pi)^r}.$$

In the case of independent ascertainments from many sources, the probability that a proband has t ascertainments, when a truncated Poisson distribution

Table 2

Analysis of Cystic Fibrosis Data

Approach Used and Information Considered	p	x
Bayesian:[a]		
SR	.25 (.20, .33)	.02 (0, .20)
SR + RA	.26 (.22, .31)	.04 (0, .17)
SR + T	.27 (.23, .32)	.03 (0, .14)
SR + RA + T	.26 (.22, .31)	.03 (0, .15)
Classical:[b]		
SR + RA + T	.26 (.23, .29)	−.02 (−.11, .07)

[a] Modal values and 95% credible intervals (in parentheses) for the univariate PMD of p and x.
[b] Joint estimates and asymptotic 95% confidence interval (in parentheses).

(Morton 1959) is assumed, is

$$Pr\{t|\pi\} = \frac{[-\ln(1-\pi)]^t(1-\pi)}{t!\pi} .$$

Making Inferences

Inferences about the parameters of interest can be made through point estimates and through the measure of uncertainty of the point estimates as evaluated on the basis of the PMD. The point estimate is the modal value of the PMD and can be obtained by finding the argument for the maximum of the distribution. The measure of uncertainty is the credible region (or highest-density region), the Bayesian counterpart for the confidence region (Box and Tiao 1973). The credible region of level α is the smallest subset Θ_α of the parameter space Θ such that, in the PMD, the event $\theta \in \Theta_\alpha$ has probability α. In spite of the analogy between the credible region and the confidence interval of the classical statistics, they are conceptually different. A credible region α means that the true value of the parameter has a probability α of belonging to the set Θ_α. On the other hand, a confidence interval α means that, if the construction method has been applied to all possible samples, $100\alpha\%$ of the confidence intervals thus calculated should contain the true value of the parameter. In the univariate case the credible region is called the credible interval. General definitions, properties, and algorithms for evaluating the credible regions can be found in Pereira and Rogatko (1984) and Rogatko et al. (1986).

Results

Cystic Fibrosis

Table 2 shows the results of the Bayesian and classical approaches as applied to the cystic fibrosis data. To make the comparison easier, only the univariate PMD are evaluated for p and x in the Bayesian approach. The modal values and the 95% credible intervals are calculated in the Bayesian approach, and the unconstrained joint maximum-likelihood estimates and the asymptotic 95% confidence intervals are evaluated in the classical approach. In the first line of table 2 (information from the SR table only), uniform priors are assigned for p, x, and π. Informative prior distributions are assigned for π according to the type of information incorporated (RA, T, or RA and T tables) from the second to the fourth lines; uniform priors are assigned for p and x. A truncated binomial distribution model is considered in the case of the RA table, and a truncated Poisson distribution model is utilized for the T table. The total amplitude of the credible or confidence interval for p and x, obtained by adding the individual amplitudes of p and x, is greater if no information about π is provided (SR table alone). It is worth noting that the total amplitude when one uses all the information is the same with the two approaches. When the Bayesian approach is used with these particular data, the modal values seem to be insensitive to the amount of information incorporated in the prior distribution for π.

Simulated Data

Deterministic samples are generated with various values of π. The univariate estimates of p and x when one assumes an unknown π are shown in table 3. It can be seen that $\hat{p}$ is overestimated when $\pi \to 0$ and is underestimated when $\pi \to 1$. Conversely, x is overestimated when $\pi \to 1$ and probably is underestimated when $\pi \to 0$. However, the hypotheses of $p = .25$ and $x = 0$ were accepted in all cases by the 95% credible-interval inclusion criterion.

Table 3

Estimation of p and x through Deterministic Samples when π Is Assumed to Be Unknown

	True π				
	.0001	.2	.5	.8	1.0
N = 100:					
$\hat{p}$	.30	.28	.26	.23	.23
$\hat{x}$	.00	.00	.01	.04	.11
N = 500:					
$\hat{p}$	.28	.26	.25	.23	.23
$\hat{x}$	.00	.00	.00	.04	.12

NOTE.—True p = .25; true x = 0; s = 5.

To study the consistency of the estimator for p, deterministic samples were generated, fixing $x = 0$, for various values of π, s, and N. Table 4 shows that for small N (e.g., $N = 50$), p is underestimated if $\pi \rightarrow 1$ and overestimated if $\pi \rightarrow 0$, whatever s. For large samples it is shown that s plays an important role, a finding that agrees with the intuition. It can be shown empirically that for $s = 2$ the estimator of p is inconsistent. For $s > 5$, however, it seems that p is consistent. Rigorous analytical proofs are still under investigation.

An evaluation of the effective type I error for some representative cases is displayed in table 5. Pseudorandom samples were generated fixing $p = .25$ and $x = 0$, and varying π and the sample size. One hundred samples were generated for each pair of π and N. The univariate PMDs of p and x were evaluated assuming π to be unknown. Mode and 95% credible intervals were calculated, and the hypotheses of $p = .25$ and $x = 0$ were tested using the 95% credible-interval inclusion criterion. The frequency of acceptance of both hypotheses $p = .25$ and $x = 0$ is higher when π is close to .5, lower for $p = .25$ when $\pi \rightarrow 0$, and lower for $x = 0$ when $\pi \rightarrow 1$.

The effective type II error can be evaluated using tables 6 and 7. When the same decision procedure described previously was used, 100 pseudorandom samples of size 100 were generated for each pair of true π and true p (table 6) and for each of true π and true x (table 7). Table 6 shows the over- and underestimation of p for low and high values of π, respectively. The frequency of type II errors, β, decreases faster from the maximum value (which varies with π) if p increases above the maximum than if p decreases below the maximum. For example, for true $\pi = .8$, the maximum β is given when the true $p = .25$, i.e.,

Table 4

Estimation of p through Deterministic Sampling, When x Is Assumed to Be Zero and π Is Assumed to Be Unknown, for Various Values of Sibship Size (s) and Sample Size (N)

	π				
s AND N	.0001	.2	.5	.8	1.0
2:					
50	.30	.30	.25	.20	.18
100	.31	.28	.25	.21	.18
500	.29	.26	.23	.20	.17
1,000	.28	.26	.23	.19	.16
2,000	.27	.25	.22	.18	.16
3,000	.26	.25	.22	.18	.16
4,000	.26	.25	.22	.18	.16
3:					
50	.30	.29	.25	.22	.18
100	.30	.27	.24	.21	.19
500	.28	.26	.23	.20	.18
1,000	.27	.25	.22	.19	.23
2,000	.27	.25	.22	.24	.23
3,000	.26	.25	.22	.24	.24
4,000	.26	.24	.21	.25	.24
4:					
50	.31	.29	.24	.21	.18
100	.30	.27	.25	.22	.20
500	.28	.26	.23	.23	.23
1,000	.27	.25	.22	.25	.24
2,000	.27	.25	.22	.25	.24
3,000	.26	.24	.25	.25	.24
4,000	.26	.24	.24	.25	.24
5:					
50	.29	.28	.24	.21	.20
100	.29	.27	.24	.21	.21
500	.27	.25	.23	.25	.24
1,000	.27	.25	.24	.25	.24
2,000	.27	.25	.25	.25	.24
3,000	.26	.24	.25	.25	.24
4,000	.26	.24	.25	.25	.25
6:					
50	.30	.29	.24	.21	.20
100	.29	.27	.26	.23	.23
500	.27	.25	.24	.25	.24
1,000	.27	.25	.24	.25	.24
2,000	.26	.25	.25	.25	.25
3,000	.26	.24	.25	.25	.25
4,000	.26	.25	.25	.25	.25
7:					
50	.30	.27	.24	.23	.21
100	.28	.27	.25	.22	.22
500	.27	.26	.25	.25	.24
1,000	.27	.25	.25	.25	.24
2,000	.26	.24	.25	.25	.25
3,000	.26	.24	.25	.25	.25
4,000	.26	.26	.25	.25	.25
8:					
50	.30	.27	.25	.22	.21
100	.29	.28	.25	.23	.23
500	.29	.25	.25	.25	.24
1,000	.27	.25	.25	.25	.24
2,000	.27	.26	.25	.25	.25
3,000	.26	.25	.25	.25	.25
4,000	.26	.25	.25	.25	.25

NOTE.—True p = .25.

Table 5

Frequency of Acceptance of the Hypotheses $p = .25$ and $x = 0$, When π Is Assumed to Be Unknown, in 100 Trials for Each Pair of Values of True π and N

True π and N	Frequency of Acceptance (%)	
	p	x
.001:		
50	68	96
100	67	97
500	65	99
.2:		
50	92	95
100	93	100
500	97	97
.5:		
50	96	96
100	99	99
500	100	98
.8:		
50	97	92
100	98	90
500	96	91
1.0:		
50	98	83
100	99	80
500	82	75

NOTE.—True $p = .25$; true $x = 0$.

$\beta_{.8}(.25) = .98$ is maximum for $\pi = .8$; and $\beta_{.8}(.25 - \delta) < \beta_{.8}(.25 + \delta)$, for $\delta = .05, .1, .15$. Table 7 shows the overestimation of x for high values of π. When tables 5 and 6 are compared, for the same π, the rate that β decreases for x is smaller than the rate that it increases for p when they vary by an equivalent amount from the maximum. This was expected, since the information about x, for a given sample size, is smaller than that for p. The last line of table 7 displays the β's for $x = .40$, but with a smaller sample size, $N = 30$, comparable with the sample size of the DEB^- group in the FA data.

Fanconi Anemia

The joint and the univariate PMD for the three groups of FA cases are obtained by assuming uniform prior distributions for p, x, and π.

Figure 1 shows the joint PMD of p and x, the modal values, and the 95% credible region for the three samples. The modal value (p, x) for DEB^+ is (.24, .00); for DEB^- it is (.23, .21); and for LITERATURE it is (.29, .13). The credible regions of the three samples include the point (.25, .00), which defines a monogenic autosomal recessive mode of inheritance.

Figures 2 and 3 show the univariate PMDs (standardized likelihood) of p and x, respectively. The modal values and the 95% credible intervals (in parentheses) are also indicated. The credible intervals of the three samples include the point $p = .25$ (fig. 2) and $x = 0$ (fig. 3).

Therefore, p and x are not significantly different from .25 and 0, respectively, in the three samples, which can thus be considered as homogeneous with respect to these two parameters. However, the modal values suggest that x might be greater in the DEB^- group, intermediate in the LITERATURE group, and zero only in the DEB^+ group. As it was shown in table 7, the probability of accepting the hypothesis x

Table 6

Frequency of Acceptance of the Hypotheses $p = .25$ and $x = 0$, When π Is Assumed to Be Unknown, in 100 Trials for Each Pair of Values of True π and True p

True p	Frequency of Acceptance at True π of (%)									
	.001		.2		.5		.8		1.0	
	p	x	p	x	p	x	p	x	p	x
.10	28	90	27	89	13	84	19	75	25	56
.12	51	90	41	89	28	85	26	78	26	60
.15	90	94	80	95	51	92	41	85	62	85
.20	100	99	100	99	96	94	80	70	72	78
.25	67	97	93	100	99	99	98	90	99	80
.30	6	95	23	97	99	99	90	92	95	84
.35	0	94	1	97	14	92	38	92	62	85
.40	0	99	0	96	0	89	3	91	7	91

NOTE.—True $x = 0$; $s = 5$; $N = 100$.

Table 7

Frequency of Acceptance of the Hypotheses $p = .25$ and $x = 0$ When π Is Assumed to Be Unknown, in 100 Trials for Each Pair of Values of True π and True x

	Frequency of Acceptance at True π of (%)									
	.001		.2		.5		.8		1.0	
True x	p	x	p	x	p	x	p	x	p	x
.00	67	97	93	100	99	99	98	90	99	80
.05	83	93	91	97	99	92	98	88	91	62
.10	89	89	97	92	98	82	98	69	95	61
.15	84	71	94	78	98	67	92	68	93	43
.20	81	62	98	66	95	59	93	41	95	35
.25	86	48	94	50	95	36	94	37	89	24
.30	90	40	96	24	96	24	88	22	92	9
.35	89	48	93	28	98	21	87	17	88	9
.40	91	12	97	9	95	12	88	16	85	5
.40	86	48	99	40	98	19	99	16	99	11

NOTE.—True $p = .25$; $s = 5$; $N = 100$, except for the last row, in which $N = 30$.

$= 0$ when $x \neq 0$ is high. For true $x = .15$ and $N = 100$ (comparable with the LITERATURE values), it is 43%–78%; and for true $x = .40$ and $N = 30$ (comparable with the DEB$^-$ values), it is 11%–48%. Thus, on the basis of the modal values, (*a*) DEB$^+$ follows a monogenic autosomal recessive mode of inheritance, (*b*) DEB$^-$ likely is a heterogeneous group, consisting of genetic and nongenetic entities, and (*c*) LITERATURE can be interpreted as a mixture of DEB$^+$ and DEB$^-$ groups.

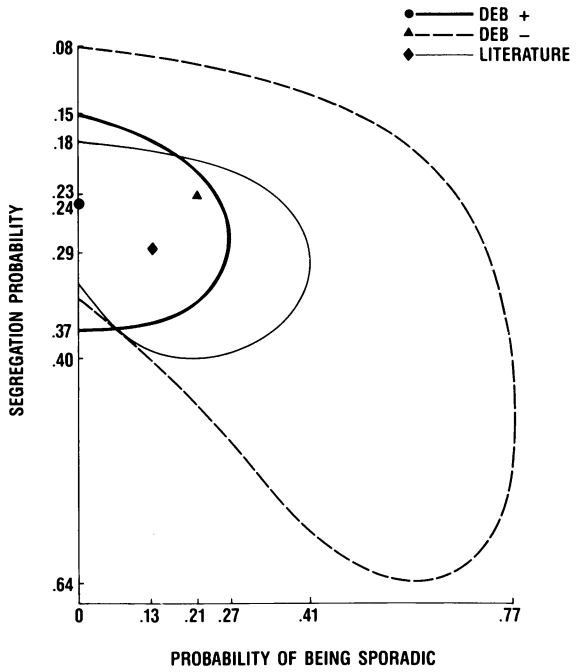

Figure 1 Joint maximum-likelihood estimates of the segregation probability, probability of being sporadic, and 95% credible region of the joint PMD for DEB$^+$ (●, —), DEB$^-$ (▲, ---), and LITERATURE (◆, —).

Discussion

The analysis of the conditional distribution of p and x for a given π is a reasonable approach when

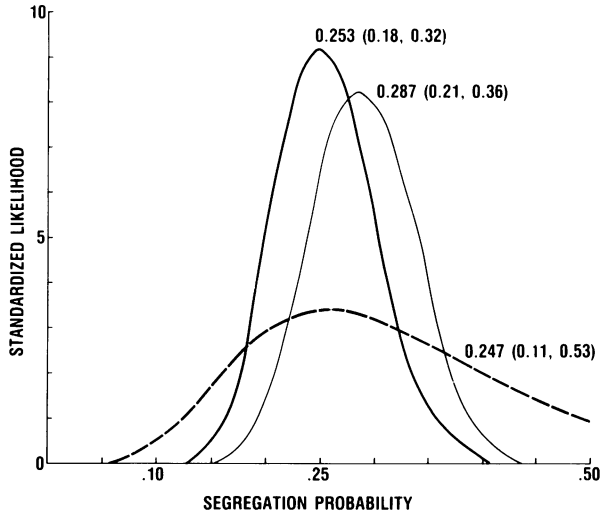

Figure 2 Standardized PMD of the segregation probability for DEB$^+$ (—), DEB$^-$ (---), and LITERATURE (—). Maximum-likelihood estimates and 95% credible intervals (in parentheses) are indicated for each curve.

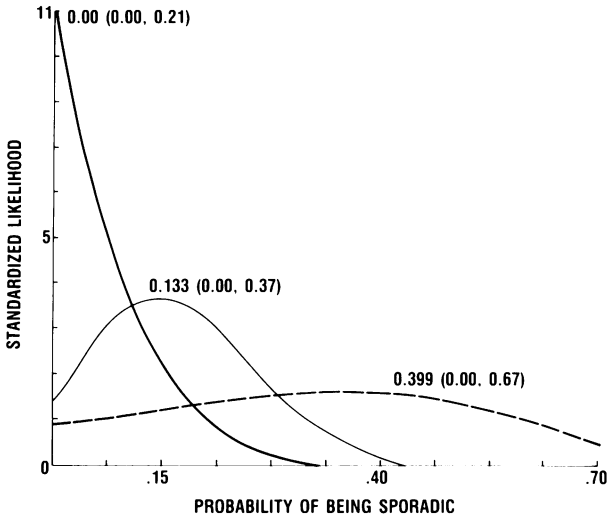

Figure 3 Standardized PMD of the probability of being sporadic for DEB^+ (—), DEB^- (---), and LITERATURE (—). Maximum-likelihood estimates and 95% credible intervals (in parentheses) are indicated for each curve.

the uncertainty about the ascertainment is small. However, when the information about the ascertainment is poor, the conditional approach is inadvisable because it assumes a precise knowledge of π. It is the equivalent, in a Bayesian approach, of concentrating all the mass of the prior distribution in a single point.

Stene (1981) showed that π is ancillary to the distribution of affected individuals (formula [1]) and that maximizing p, x, and π together has no statistical support. Ewens and Shute (1986*b*) proposed "a resolution of the ascertainment sampling problem" when it is possible to divide the genetic information into two parts, one that is relevant to ascertainment and one that is not. In the case of rare recessive disorders, all the genetic information is related to an ascertainment scheme; thus their method cannot be used.

Stene (1975) suggested a more general model for the ascertainment, π^α, $\alpha > 0$. Ewens and Shute (1986*b*) generated examples of quadratic ascertainment ($\alpha = 2$), which can happen when a two-stage selection of the families occurs. We considered the usual model of $\alpha = 1$ for the cystic fibrosis and FA data, since there was no indication of multistage selection of the families. In other instances it might be necessary to include α in the model, a decision that will result in the need to eliminate another nuisance parameter.

In the Bayesian approach, the elimination of π by integrating it out may be interpreted as the evaluation of the expected distribution of the parameters of interest weighted by the prior distribution. We have used uniform prior distributions in every case in this paper; that is, we have considered the information contained in the likelihood function only. However, when the researcher has an idea about the ascertainment, it can be translated into a suitable prior distribution.

The Bayesian approach for eliminating π is a valid statistical solution which provided reliable results. In the cystic fibrosis data the estimated values of p and x did not vary with the amount of information considered for π. In the simulated data, the recovery of the true values of p and x were within an acceptable level of tolerance. Convergence problems did not occur in any case. On the other hand, multiple integration may be time consuming, but fast machines are available and the progress in computing speed is impressive. Also, parallel computers will greatly reduce the computing time required for multiple integrations, since algorithms for integration can be structured in a parallel fashion. Today algorithms and implementations are available to integrate as many as 20 variables simultaneously (Naylor and Shaw 1985). Therefore the procedure advocated here may be used in more complex models of segregation analysis to avoid the frequent problems of convergence that are due to flat likelihood surfaces.

Although the method did perform well when a noninformative uniform prior distribution for π was used, it is still preferable to design carefully the data collection and to gather information on π. Since π is ancillary to p, the lack of information about π will always increase the uncertainty in p. However, when the ascertainment scheme is uncertain and when the assumptions of independent and constant ascertainment are reasonable, we claim that the method here presented is the best valid statistical procedure that can be used.

Acknowledgments

This work was supported in part by U.S. Public Health Service grants GM36295 (to A.R.) and HL32987 (to A.D.A.) from the National Institutes of Health, Basil O'Connor Starter Research Grant 5-446 from the March of Dimes Birth Defects Foundation (to A.D.A.), General Clinical Research Center grant RR00102 from the National Institutes of Health to The Rockefeller University Hospital, and support from the Pew Memorial Trust to the Laboratory for Investigative Dermatology.

References

Alter, B. P., and N. U. Potter. 1983. Long-term outcome in Fanconi's anemia: description of 26 cases and review of the literature. Pp. 43–62 *in* J. German, ed. Chromosome mutation and neoplasia. Alan R. Liss, New York.

Auerbach, A. D., B. Adler, and R. S. K. Changanti. 1981. Prenatal and postnatal diagnosis and carrier detection of Fanconi anemia by a cytogenetic method. Pediatrics **67**:128–135.

Auerbach, A. D., A. Rogatko, and T. Schroeder. 1988. International Fanconi Anemia Registry (IFAR): first report. *In* T. M. Schroeder, A. D. Auerbach, and G. Obe, eds. Clinical and experimental aspects of Fanconi anemia. Springer, Heidelberg (in press).

Auerbach, A. D., M. Sagi, and B. Adler. 1985. Fanconi anemia: prenatal diagnosis in 30 fetuses at risk. Pediatrics **76**:794–800.

Auerbach, A. D., and S. R. Wolman. 1976. Susceptibility of Fanconi's anemia fibroblasts to chromosome damage by carcinogens. Nature **261**:494–496.

Auerbach, A. D., M. Zhang, R. Ghosh, E. Pergament, Y. Verlinsky, H. Nicholas, and J. Boue. 1986. Clastogen-induced chromosomal breakage as a marker for first trimester prenatal diagnosis of Fanconi anemia. Hum. Genet. **73**:86–88.

Basu, D. 1977. On the elimination of nuisance parameters. J. Am. Stat. Assoc. **72**:355–366.

Boor, C. 1971. CADRE: an algorithm for numerical quadrature. Pp. 32–35 *in* J. R. Rice, ed. Mathematical software. Academic Press, New York.

Box, G. E. P., and G. C. Tiao. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, MA.

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

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

Ewens, W. J., and N. C. E. Shute. 1986*a*. The limits of ascertainment. Ann. Hum. Genet. **50**:399–402.

———. 1986*b*. A resolution of the ascertainment sampling problem. I. Theory. Theor. Popul. Biol. **30**:388–412.

Fanconi, G. 1967. Familial constitutional panmyelocytopathy, Fanconi's anemia (F.A.). I. Clinical aspects. Sem. Hematol. **4**:233–240.

Fisher, R. A. 1934. The effect of methods of ascertainment upon the estimation of frequencies. Ann. Eugenics **6**:13–25.

Glanz, A., and F. C. Fraser. 1982. Spectrum of anomalies in Fanconi anemia. J. Med. Genet. **19**:412–416.

Greenberg, D. A. 1986. The effect of proband designation on segregation analysis. Am. J. Hum. Genet. **39**:329–339.

Haldane, J. B. S. 1938. The estimation of the frequencies of recessive conditions in man. Ann. Eugenics **14**:33–341.

IMSL library. 1982. IMSL, Inc., Houston.

Morton, N. E. 1959. Genetic tests under incomplete ascertainment. Am. J. Hum. Genet. **11**:1–16.

Naylor, J. C., and J. E. H. Shaw. 1985. Bayes four—user guide. Tech. rep. 9-85. Department of Mathematics, University of Nottingham, Nottingham.

Pereira, C. A. B., and A. Rogatko. 1984. The Hardy-Weinberg equilibrium under a Bayesian perspective. Braz. J. Genet. **7**:689–707.

Rogatko, A., C. A. B. Pereira, and O. Frota-Pessoa. 1986. A Bayesian method for the estimation of penetrance: application to mandibulofacial and fronto-nasal dysostoses. Am. J. Med. Genet. **24**:231–246.

Schroeder, T. M., F. Anschultz, and A. Knoff. 1964. Spontane Chromosomenaberrationen bei familiärer Panmyelopathie. Humangenetik **1**:194–196.

Schroeder, T. M., D. Tilgen, J. Kruger, and F. Vogel. 1976. Formal genetics of Fanconi's anemia. Hum. Genet. **32**:257–288.

Stene, J. 1975. Sampling methods in human genetics with applications to Down's syndrome. Ph.D. thesis, University of Copenhagen, Copenhagen.

———. 1977. Assumptions for different ascertainment models in human genetics. Biometrics **33**:523–527.

———. 1981. Probability distributions arising from the ascertainment and the analysis of data on human families and other groups. Pp. 233–264 *in* C. Taillie, G. P. Patil, and B. A. Baldessari, eds. Statistical distributions in scientific work. Vol. **6**. D. Reidel, Dordrecht.

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