AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 70:489–503 (1986)

Frequencies of Complex Diseases in Hybrid Populations

RANAJIT CHAKRABORTY AND KENNETH M. WEISS
Center for Demographic and Population Genetics, University of Texas Graduate School of Biomedical Sciences, Houston, Texas 77225 (R.C.); Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania 16802 (K.M.W.)

KEY WORDS Admixed populations, Complex etiology, Mexican-Americans, Type II diabetes mellitus

ABSTRACT Diseases of complex etiology demonstrate considerable variation in their frequencies in different ethnic populations. Noninsulin-dependent diabetes mellitus (NIDDM), rheumatoid arthritis, and several cardiovascular diseases constitute examples of such disorders. In genetic studies involving hybrid populations of known ancestry, it is of interest to compare and correlate disease prevalence with the admixture proportion, the latter estimated from a number of polymorphic genetic markers. Theoretical formulations are provided relating disease prevalence in a hybrid population to the admixture proportion under different models of disease transmission. It is shown that the relationship between admixture proportion and disease frequency provides discriminatory power regarding the mode of inheritance. This method is illustrated with an example comparing the proportion of Amerindian ancestry in Mexican-Americans and the prevalence of NIDDM. It is found that genetic factors are involved in susceptibility to NIDDM, but the mode of inheritance cannot be explained by any simple genetic model, and the role of sporadic events cannot be totally ruled out.

Populations derived from a common ancestral stock exhibit similarities in their allele frequency structure (Mourant et al., 1976). A similar phenomenon is also seen in the frequencies of diseases of simple mendelian inheritance (Ramot, 1974; Wagener and Cavalli-Sforza, 1975; Chakravarti and Chakraborty, 1978; Wagener et al., 1978). This is true in spite of local fluctuations in selection intensity that sometimes perturb clinal relationships between the incidence of genetic disorder and genetic affinity of populations. Such findings have been the basis for epidemiological studies of aggregation at the population level (Haenszel, 1982; Lilienfeld and Lilienfeld, 1982; Chakraborty and Szathmary, 1985), indicating the significance of genetic factors in the etiology of chronic disorders. Nevertheless, it is often not possible to determine the exact mode of transmission solely from population association studies.

Nevertheless, several interesting studies on the effect of migration on the occurrence of chronic diseases have been carried out to understand the role of environmental factors on such diseases. Steiner (1954) and Haenszel (1982) provide some review on such materials in the context of geographic pathology of cancer. However, most epidemiologists have used migration in a rather narrow sense, referring to the occurrence of disease in the migrants themselves or in their direct and immediate descendents. These studies, in general, examined the effects of changed environments on essentially unchanged genotypes (to the extent that a group of migrants represent a gene pool representative of their homeland). It is often difficult to distinguish between genetic effects and environmental ones in these studies, because migrants and their families tend to preserve at least some of their cultural practices, diet and religion included.

In the case of genetic epidemiology, it is the genes themselves and not the bearers of those genes that are of interest. Thus, it may

Received January 8, 1986; revision accepted April 2, 1986.

be informative to examine populations with different amounts of genetic admixture between source populations with very different disease risks. To the extent that there is a systematic relationship between admixture and disease occurrence, especially if this relationship resembles theoretical predictions, one may be able to infer that genes, and not environments, are mainly involved in causation of disease.

Furthermore, since almost all contemporary populations are products of past generations of gene flow, the purpose of this presentation is to examine the relationship between the disease prevalence and the degree of admixture in a population of known ancestral composition. For diseases with simple mendelian or polygenic modes of inheritance, we show that the relationship between disease frequency and the accumulated level of admixture is quite simple and that the relationship provides some discriminatory power for the different modes of inheritance. Since low expressivity of genes, occurrences of sporadic cases, and/or environmental factors often disturb this relationship for complex disorders in an uncertain fashion, we argue that departures from the predictions of these models should provide indicators of the relative importance of nongenetic etiological factors. It is evident that reliable knowledge of accumulated admixture in a population is essential for deployment of these analytic strategies. With data on type II diabetes diabetes mellitus (NIDDM) prevalence in populations of varying degrees of Amerindian admixture, we show that while the relationship of prevalence to percent Amerindian ancestry does not quite agree with the prediction from any particular mendelian model, nevertheless it is clear that genetic factors are quite important in the predisposition of NIDDM.

THE RELATIONSHIP BETWEEN ADMIXTURE AND DISEASE PREVALENCE

Let H be a hybrid population that consists of a fraction (m) of its gene pool drawn from population I and a fraction (1 − m) drawn from ancestral population II. For a two-allelic locus with genotypes AA, Aa, and aa, the expected genotypic frequencies in the hybrid population (after some generations of random mating) are given by

$$\text{freq. (AA)} = p_H^2 = [mp_1 + (1 - m)p_2]^2 \quad (1a)$$

$$\text{freq. (Aa)} = 2p_Hq_H = 2[mp_1 + (1 - m)p_2][mq_1 + (1 - m)q_2] \quad (1b)$$

and

$$\text{freq. (aa)} = [mq_1 + (1 - m)q_2]^2, \quad (1c)$$

where p_1, p_2, and p_H are the frequencies of the A allele in populations I, II, and H, respectively, and q is the frequency of the a allele in the respective populations.

Dominant disease

The frequency of a dominant genetic disease is given by

$$d_i = p_i^2 + 2p_iq_i = 1 - q_i^2, \quad (2)$$

for i = I, II, and H.

Using equations 1a – 1c and 2, we may rewrite the frequency of the disease in the hybrid population as

$$\begin{aligned} d_H &= 1 - q_H^2 = 1 - [mq_1 + (1 - m)q_2]^2 \\ &= 2m(1 - m)[1 - \sqrt{(1 - d_1)(1 - d_2)}] \\ &\quad + m^2d_1 + (1 - m)^2d_2 \end{aligned} \quad (3a)$$

Recessive disease

The frequency of a recessive disease in the i-th population, on the other hand, is $d_i = q_i^2$ for i = I, II, and H; and the corresponding relationship becomes

$$\begin{aligned} d_H &= q_H^2 = [mq_1 + (1 - m)\,q_2]^2 \\ &= m^2d_1 + (1 - m)^2d_2 \\ &\quad + 2m(1 - m)\sqrt{d_1d_2} \end{aligned} \quad (3b)$$

In the simple case where the disease allele is fixed in population I but completely absent in population II ($d_1 = 1$, $d_2 = 0$), equations 3a and 3b reduce to

$$d_H = 1 - (1 - m)^2, \quad \text{for a dominant disease,} \quad (4a)$$

and

$$d_H = m^2, \text{ for a recessive disease.} \quad (4b)$$

Polygenic threshold disease

A polygenic threshold disease is usually modeled by a threshold treatment (Critten-

den, 1961; Falconer, 1965), a model that assumes a liability factor, X, consisting of the sum total of two additive components, G and E. The genotypic value G is composed of additive effects of a large number of loci, each with small impact, and the environmental value E is composed of a large number of small, additive environmental contributions. We assume that E is normally distributed with mean $\mu_E = 0$ and variance σ_e^2. Specification of the distribution of G depends on the number of loci underlying disease liability, the allele frequencies at each locus, and their allelic effect. If we assume that there are two alleles (A_j and a_j) at the j-th locus and that the allelic effect of A_j is α_j and that of a_j is zero, then in the absence of dominance effects, traditional quantitative genetic theory (Falconer, 1960) yields

$$\text{Mean of G, } \mu_g = 2 \sum_j \alpha_j p_j \qquad (5a)$$

and

$$\text{Variance of G, } \sigma_g^2 = 2 \sum_j \alpha_j^2 p_j (1 - p_j), \qquad (5b)$$

where the Σ is over all loci (say, N).

In a particular population, since $\mu_E = 0$, the distribution of liability may be approximated by a normal distribution with

$$\text{Mean} = \mu_X = 2 \sum_j \alpha_j p_j \qquad (6a)$$

and

$$\text{Variance} = \sigma_X = \sigma_g^2 + \sigma_e^2 = 2 \sum_j \alpha_j^2 p_j (1 - p_j) + \sigma_e^2, \qquad (6b)$$

assuming that the number of loci is large and that G and E are independently distributed (no gene–environment interaction).

If t denotes the threshold of liability, beyond which an individual becomes affected, the disease frequency becomes

$$\begin{aligned} d &= \Pr.(X \geq t) \\ &= \int_t^\infty \frac{1}{\sqrt{2\pi(\sigma_g^2 + \sigma_e^2)}} \exp \left[-\frac{(x - \mu_g)^2}{2(\sigma_g^2 + \sigma_e^2)} \right] dx \\ &= 1 - \Phi \left[\frac{t - \mu_g}{\sqrt{(\sigma_g^2 + \sigma_e^2)}} \right], \end{aligned} \qquad (7)$$

where $\Phi(z)$ is the cumulative distribution function for a standard normal variable, $z \sim N(0,1)$.

While μ_g for a hybrid population may be written in terms of admixture components and the corresponding means of populations I and II,

$$\mu_g(H) = m\mu_g(1) + (1 - m)\mu_g(2), \qquad (8)$$

no such simple relation exists between $\sigma_g^2(H)$, $\sigma_g^2(1)$, and $\sigma_g^2(2)$. Thus, in the general case, the frequency of polygenic threshold disease (d_H) cannot be written as a simple function of m, d_1, and d_2 alone, except for the cases of simple dominant and recessive diseases. This is intuitively clear, since an assumption that $\sigma_e^2 = 0$ and assigns the threshold value to either the heterozygote or one of the homozygous genotypic classes.

For the polygenic model, using the restrictive assumptions that $\sigma_e^2 = 0$ and that allele frequencies and allelic effects at each locus are the same ($p_j = p$, $\alpha_j = \alpha$), we obtain

$$\begin{aligned} &\mu_g = 2N \alpha p, \\ \text{and} \quad &\sigma_g^2 = 2N \alpha^2 p(1 - p). \end{aligned} \qquad (9)$$

In this case, however, X (= G + E) may not be normally distributed unless the number of loci (N) is large and p is intermediate in frequency. Note that G is the sum total of N independently distributed trinomial variables, i.e.,

$$G = \sum_{j=1}^{N} G_j, \qquad (10)$$

where

$$G_j = \begin{cases} 0 \text{ with probability } (1 - p)^2, & (10a) \\ \alpha \text{ with probability } 2p(1 - p), & (10b) \\ 2\alpha \text{ with probability } p^2. & (10c) \end{cases}$$

The probability distribution of G may be evaluated by the probability generating function (pgf):

$$\Pr.\{G = K\alpha\} = \text{Coefficient of } s^K \text{ in } [(1 - p)^2 + 2p(1 - p)s + p^2 s^2]^N \qquad (11)$$

for any arbitrary s satisfying $-1 \leq s \leq 1$ (Feller, 1967), for K = 0 to 2N.

Alternatively,

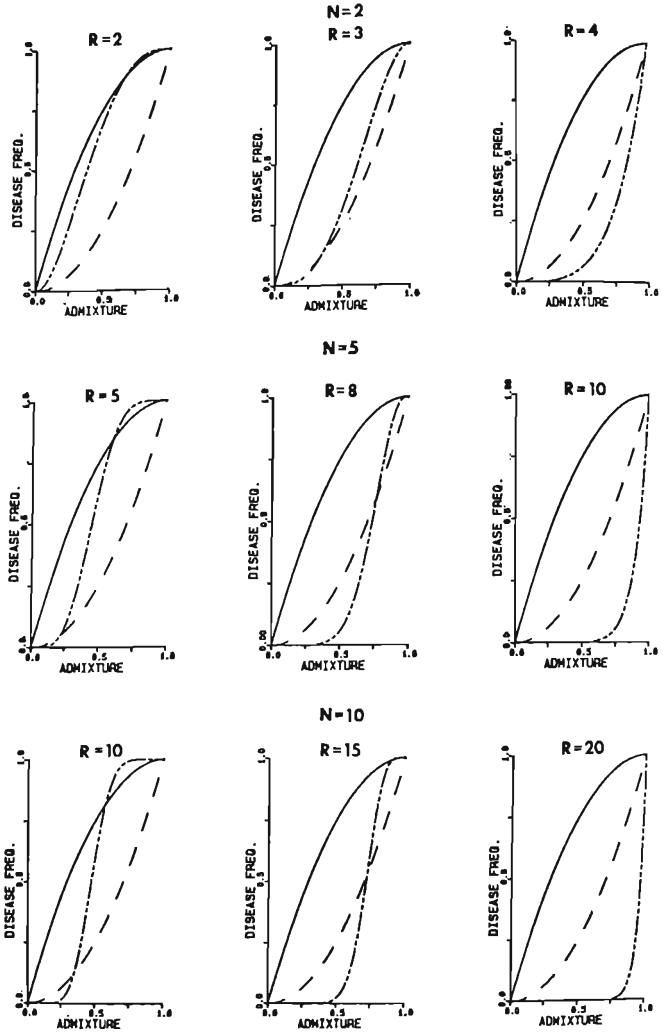

Fig. 1. Relationship between admixture level and disease frequency in a hybrid population under different modes of inheritance: solid lines for autosomal dominant model, dashed and dotted lines are for autosomal recessive model, and dashed lines are for a polygenic threshold model. N represents the number of loci, and R is the position of the threshold for the polygenic model. In all of these computations no environmental factor is assumed to be responsible for the disease, and $d_1 = 1.0$, $d_2 = 0.0$.

$$\Pr.\{G = (2N_1 + N_2)\alpha\} = \frac{N!}{N_1!N_2!N_3!}(p^2)^{N_1}(2pq)^{N_2}(q^2)^{N_3}, \quad (12)$$

for $N_1 + N_2 + N_3 = N$.

Equation 12 yields the probability of the event $\{G \geq t\}$, for any fixed $t = R\alpha$ ($0 \leq R \leq 2N$), specifying the frequency of a disease of threshold t. Noting that the expected p values in a hybrid population can be written as $p_H = mp_1 + (1 - m)p_2$, we may numerically compute the relationship between d_1, d_2, and d_H as a function of N, t, and m.

In the hypothetical case, where all A_j-alleles are fixed in population I and are completely absent in population II (as is assumed in deriving equations 4a and 4b), the frequency of disease in the hybrid population is (with $K = 2N_1 + N_2$),

$$d_H = \sum_{N_1}\sum_{\substack{N_2 \\ K \geq R}}\sum_{N_3}\frac{N!}{N_1!N_2!N_3!}(m^2)^{N_1}[2m(1-m)]^{N_2}[(1-m)^2]^{N_3} \quad (13)$$

$$= \sum_{K=R}^{2N}\binom{2N}{K}m^K(1-m)^{2N-K}$$

Note that even in this simple case, the frequency of a polygenic threshold disease depends on the number of loci (N) and the threshold ($t = R\alpha$), in addition to the amount of admixture (m) in the hybrid population.

Figure 1 presents some numerical comparisons of the predictions of equations 4a, 4b, and 13. In these computations, we have used $\alpha = 1$, and $N = 2$, 5, and 10 for the polygenic model. The threshold values (t) were chosen to represent approximately the multilocus dominant, the codominant, and the recessive cases, respectively. It is clear from these illustrations that in case where the disease allele is fixed in population I, and population II is completely disease free, the frequency of the disease in the hybrid population, for each mode of transmission, rises with the contribution of population I. Although the disease frequency under dominant inheritance is always higher than for a recessive disorder, given the same amount of admixture, the relationship under the polygenic threshold model depends on the position of the threshold in relation to the range of genotypic values in a population. In general, the rate of increase in the disease frequency in the latter instance, as a function of m, is slow for low values of m. As m increases, the slope gradually rises. Unless the threshold is at an extreme genotypic value (nearly recessive, approximately), the curve for polygenic inheritance crosses the other two prediction equations. Even though these computations are based on very restrictive assumptions ($d_1 = 1$, $d_2 = 0$, no environmental and/or sporadic cases), we argue that the dependence of the disease incidence on the amount of admixture offers some discriminatory power, particularly when data on admixed populations are available for an array of hybrid populations each of which has a varying degree of admixture.

Table 1 provides computations for somewhat more general cases. Here we consider nontrivial disease frequencies in the two parental populations ($0 < d_1 \neq d_2 < 1$). However, we still neglect the role of environmental factors, so that equations 3a and 3b describe the relationship between m and d_H for autosomal dominant and recessive disorders, while equation 13, with p_H substituted for m, gives the frequency of disease in the hybrid population under a polygenic threshold model. The computations for the polygenic threshold model are somewhat tricky in this case, since for given values of N, R, and d_i we first must compute the allele frequency p_i for both parental populations. This was done by iteratively solving for p in the equation

$$d_i = \sum_{K=R}^{2N}\binom{2N}{K}p_i^K q_i^{2N-K} \quad (14)$$

for given values of d_i, N, and R. Once the allele frequencies in the two parental populations were calculated, that in the hybrid population, p_H, was computed by $mp_1 + (1 - m)p_2$, which was then substituted into equation 14 to compute the frequency of the disease in the hybrid population (d_H).

The qualitative conclusions from Figure 1 hold for this general case as well. The polygenic curve rises slowly as a function of m, and some discriminatory power is still preserved for polymorphic disease allele frequencies ($0 < p_1 \neq p_2 < 1$) in the two parental stocks. The computations presented in Table 1, however, indicate that this strategy will not be fruitful in delineating the mode of inheritance of a genetic disease from data on just a single population. We need an array of hybrid populations, all of which are

TABLE 1. *Relationship between admixture level and incidence of disease for different models of transmission in an admixed population for various values of disease prevalence in the parental populations*

m	One locus		Polygenic									
			N = 2			N = 5			N = 10			
	Dom.	Rec.	R = 2	R = 3	R = 4	R = 5	R = 8	R = 10	R = 10	R = 15	R = 20
$d_1 = 0.9, d_2 = 0.1$											
0.2	0.324	0.196	0.262	0.213	0.173	0.233	0.197	0.162	0.226	0.204	0.168
0.4	0.516	0.324	0.451	0.367	0.279	0.414	0.341	0.255	0.405	0.359	0.248
0.6	0.676	0.484	0.633	0.550	0.429	0.609	0.528	0.396	0.604	0.554	0.385
0.8	0.804	0.676	0.787	0.738	0.633	0.780	0.730	0.602	0.778	0.750	0.591
$d_1 = 0.3, d_2 = 0.01$											
0.2	0.072	0.036	0.041	0.032	0.026	0.030	0.025	0.022	0.027	0.024	0.021
0.4	0.132	0.078	0.090	0.070	0.056	0.066	0.055	0.045	0.060	0.053	0.042
0.6	0.190	0.136	0.151	0.128	0.106	0.124	0.106	0.088	0.115	0.104	0.082
0.8	0.246	0.210	0.222	0.205	0.184	0.203	0.186	0.166	0.195	0.185	0.159

derived from the same two parental stocks, with varying degrees of admixture. Since the trend in the predictions is critically dependent on the admixture level (m), to infer the role of genetic factors in chronic diseases from epidemiological studies on migrants an accurate determination of the admixture rate in migrant populations is needed.

MODELS WITH SPORADIC OCCURRENCES OF DISEASE

For almost all chronic diseases, there is a certain fraction of cases that occur sporadically. For our proposes, let c be the ratio of sporadic to genetic cases. Under this supposition, the disease frequency may be expressed as

$$\text{Disease Frequency: } d = (1 + c) \times \text{Genetic Risk.} \quad (15)$$

Dominant disease

The frequency of an autosomal dominant disease with a fraction of sporadics will be given by

$$d_i = (1 + c) \cdot (1 - q_i^2) \quad (16a)$$

or

$$q_i^2 = 1 - [d_i / (1 + c)]. \quad (16b)$$

With the same notation as before, the disease frequency in a hybrid population with a fraction m of its gene pool from population I (with disease frequency d_1), and a fraction $(1 - m)$ from population II (with disease frequency d_2), is then

$$\begin{aligned} d_H &= (1 + c) \cdot (1 - q_H^2) \\ &= (1 + c) \cdot [1 - m^2 q_1^2 - (1 - m)^2 q_2^2 \\ &\quad - 2m(1 - m) q_1 q_2] \\ &= m^2 d_1 + (1 - m)^2 d_2 + 2m(1 - m)(1 + c) \\ &\quad - \sqrt{[1 + c - d_1][1 + c - d_2]}. \end{aligned} \quad (17)$$

Note that for c = 0 (no sporadic cases), equation 17 reduces to equation 3a, as should be the case. It is interesting to note that for the same d_1 and d_2, the relationship between m and d_H rises at a slower pace as the relative proportion (c) of sporadic to genetic cases increases, as shown in Figure 2. The curve for a dominant mode of inheritance is always bounded below by the curve for the recessive model, since the term

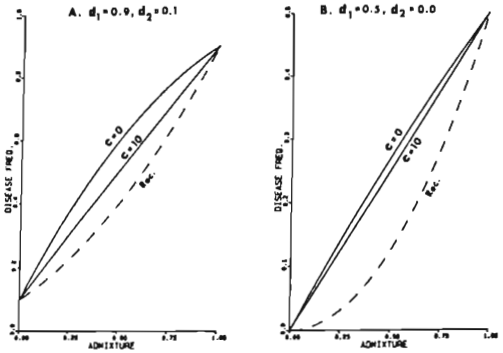

Fig. 2. Effect of sporadic occurrences of a disease on the relationship between the admixture level and disease frequency in a hybrid population. c represents the ratio of sporadic versus genetic rate of the disease. The solid lines are for the autosomal dominant mode of inheritance, and the broken lines are for the autosomal recessive model.

$$2m(1 - m)[(1 + c) - \sqrt{(1 + c - d_1)(1 + c - d_2)}] \rightarrow 2m(1 - m)\sqrt{d_1 d_2} \quad \text{as } c \rightarrow \infty, \tag{18}$$

reducing equation 17 to the form of equation 3b.

Recessive disease

Autosomal recessive disorders with a fraction of sporadics behave differently. In this case, equation 15 yields

$$d_i = (1 + c) \cdot q_i^2 \tag{19a}$$

or

$$q_i^2 = d_i / (1 + c), \tag{19b}$$

so that the disease frequency in the hybrid population is

$$d_H = (1 + c) \cdot q_H^2 = m^2 d_1 + (1 - m)^2 d_2 + 2m(1 - m) \cdot \sqrt{d_1 d_2}, \tag{20}$$

which is identical to equation 3b.

The fact that sporadic occurrences do not disturb the relationship between d_H and m is intuitively easy to explain. Since the effect of the sporadic rate in a recessive case is essentially to decrease the disease allele frequency by a factor $\sqrt{(1 + c)}$, the square of the multiplier of the genetic risk dictates the population risk of the disease, as in equation 15.

Polygenic threshold disease

Sporadic occurrences under a polygenic threshold model are taken into account by setting $\sigma_e^2 \neq 0$ in the formulation described earlier. We now define the heritability (h^2) of the liability X, which has an implicit relationship with the sporadic rate (Chakraborty, 1986). Using the usual definition of heritability, we have

$$h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2) \tag{21a}$$

or

$$\sigma_e^2 = \sigma_g^2 [(1 - h^2) / h^2] = \beta \cdot \sigma_g^2, \tag{21b}$$

where $\beta = (1 - h^2) / h^2$.

As before, assuming normal distributions for X in each population, we have

$$d_i = \int_t^\infty N(X_i | \mu_i, \sigma_i^2)\, dx_i, \quad \text{where } \mu_i = 2N\alpha p_i, \quad \text{and } \sigma_i^2 = 2N\alpha^2 p_i q_i; \tag{22a}$$

in which $N(x_i|\mu_i, \sigma_i^2)$ represents the density function of a normal variate X_i with mean μ_i, and variance σ_i^2.

Equation 22a may be manipulated to yield

$$\frac{[t - 2N\alpha p_i]^2}{\sigma_e^2 + 2N\alpha^2 p_i(1 - p_i)} = [\Phi^{-1}(1 - d_i)]^2 \quad (22b)$$

or

$$(t - 2N\alpha p_i)^2 = 2N\alpha^2 p_i(1 - p_i) \cdot (1 + \beta)[\Phi^{-1}(1 - d_i)]^2 \quad (22c)$$

with the substitution $\sigma_e^2 = \beta \cdot \sigma_g^2$.

Equation 22c can be reduced to a quadratic form

$$A_i p_i^2 - B_i p_i + C_i = 0, \quad (23)$$

with

$$A_i = 4N^2\alpha^2 + 2N\alpha^2 D_i(1 + \beta),$$
$$B_i = 4N\alpha t + 2N\alpha^2 D_i(1 + \beta),$$
$$C_i = t^2,$$

with $D_i = [\Phi^{-1}(1 - d_i)]^2$, with $\Phi^{-1}(p)$ as the normal deviate for a cumulative probability p.

Using this quadratic equation, we can evaluate the allele frequencies for the two parental populations as

$$p_i = \frac{B_i \pm \sqrt{B_i^2 - 4A_iC_i}}{2A_i} \quad (24)$$

as functions of the threshold (t), the number of loci (N), the allelic effect (α), the heritability (h^2), and the disease frequency (d_i). Note that both roots of equation 24 may be within the admissible range (0, 1). Only one of them is legitimate for a given t and d, since the assumption of normality for X_i implies that

$$t > 2N\alpha p_i \quad \text{for } d_i < 0.5 \quad (25a)$$

and

$$t < 2N\alpha p_i \quad \text{for } d_i > 0.5. \quad (25b)$$

Inequalities 25a – 25b specify the allele frequency solutions $p_i < t/(2N\alpha)$ for $d_i < 0.5$ and $p_i > t/(2N\alpha)$ for $d_i > 0.5$.

Since the hybrid population has the expected allele frequency $p_H = mp_1 + (1 - m)p_2$, once p_1 and p_2 are determined from equation 23 with restriction 24, the disease frequency is given by

$$d_H = 1 - \Phi\left[\frac{t - 2N\alpha p_H}{\sqrt{\{2N\alpha^2 p_H(1 - p_H)(1 + \beta)\}}}\right] \quad (26)$$

for a general polygenic threshold disease. We have implicitly assumed in this derivation that the heritability of liability (h^2) is the same for all populations, which is approximately true for many physiological traits (Sing and Orr, 1978; Cloninger et al., 1983).

For some disorders a more restrictive assumption may be made, namely, the total phenotypic variance of X, $\sigma^2 = \sigma_g^2 + \sigma_e^2$, is constant over all populations. From equation 22a, it is easy to show that

$$d_H = 1 - \Phi[m\Phi^{-1}(1 - d_1) + (1 - m)\Phi^{-1}(1 - d_2)], \quad (27)$$

with $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ as defined in equations 7 and 23. Even though the assumption of constancy of σ^2 in all populations may be contested, the relationship 27 is particularly interesting, since it describes a simple relationship between d_H and m without involving the parameters N, t, α, and p explicitly. In this sense, this relationship parallels those under the simple mendelian models (equations 3a, 3b, and 17).

Table 2 illustrates the effect of environmental contributions on the relationship between m and d_H for a polygenic threshold disease. For illustrative purposes, we assume that the two parental populations have disease frequencies $d_1 = 0.9$ and $d_2 = 0.1$, respectively. Let m denote the proportional ancestral contribution of parental population I to the hybrid gene pool. We assume that there are two loci contributing to the disease liability (with threshold $t = 2\alpha$, 3α, and 4α, respectively, or correspondingly R = 2, 3 and 4 with $\alpha = 1$); or there are ten loci involved (with threshold $t = 10\alpha$, 15α, and 20α, respectively, or correspondingly R = 10, 15, and 20). The effect of the heritability on the relationship between m and d_H for a multifactorial threshold disease is more complex than the effect of the sporadic rate in the case of a simple mendelian disease. The numerical values in Table 2 indicate that as β increases (i.e., as h^2 decreases), the increase of d_H with increasing m becomes confused, since in such

TABLE 2. Relationship between admixture level and disease frequency of a polygenic threshold disorder for different environmental contributions [$\beta = (1 - h^2)/h^2$]

	N = 2			N = 10		
β	R = 2	R = 3	R = 4	R = 10	R = 15	R = 20
			m = 0.2			
0.0	0.262	0.213	0.173	0.226	0.204	0.158
0.5	0.257	0.192	0.140	0.229	0.192	0.129
2.0	0.282	0.206	0.152	0.237	0.188	0.132
10.0	0.348	0.264	0.203	0.270	0.199	0.146
			m = 0.4			
0.0	0.451	0.367	0.279	0.405	0.359	0.248
0.5	0.420	0.306	0.187	0.404	0.327	0.166
2.0	0.431	0.315	0.208	0.409	0.313	0.172
10.0	0.455	0.361	0.277	0.425	0.309	0.197
			m = 0.6			
0.0	0.633	0.550	0.429	0.604	0.554	0.385
0.5	0.581	0.445	0.244	0.596	0.506	0.217
2.0	0.570	0.433	0.269	0.591	0.477	0.225
10.0	0.544	0.442	0.337	0.574	0.437	0.257
			m = 0.8			
0.0	0.787	0.738	0.633	0.778	0.750	0.591
0.5	0.743	0.626	0.320	0.770	0.712	0.292
2.0	0.717	0.586	0.342	0.763	0.680	0.300
10.0	0.652	0.537	0.395	0.730	0.603	0.331

In these computations we assumed that the disease frequency in the donor population (d_1) is 0.9 and that in the receiving population (d_2) is 0.1.

cases the disease is governed more by variation in the environmental factors among individuals, which changes independently of the genes inherited.

From these computations, we conclude that if the disease is totally environmental in origin, we cannot expect any trend in the relationship between admixture and disease frequency in hybrid populations. Genetic factors responsible for the disease will necessarily show a continuous change in disease prevalence with the degree of admixture, whatever be the mode of transmission for the disease. Furthermore, although not conclusive, some discriminatory power does exist in the prediction equations, unless environmental changes in hybrid populations are exactly parallel to genetic admixture levels.

THE RELATIONSHIP BETWEEN NIDDM PREVALENCE AND ADMIXTURE IN ADMIXED POPULATIONS OF NORTH AMERICA

Noninsulin-dependent diabetes mellitus (NIDDM) is a major chronic disease whose prevalence varies greatly in different parts of the world, and there are strong suggestions that these vary with changes in the genetic make-up of populations, including that in admixed groups (e.g., Zimmet, 1982; Kirk et al., 1985; Chakraborty et al., 1986). However, the correspondence of amount of admixture and NIDDM prevalence has been tested specifically in a few reports (Stern et al., 1984; Hanis et al., 1986; Bennett, 1986), all of which compared the prevalence rates to the fractions of Amerindian genes in the populations, estimated either by the number of non-Amerindian grandparents (Bennett, 1986), from genetic marker data (Chakraborty et al., 1986), or from skin reflectance (Gardner et al., 1984).

Table 3 presents a compilation of NIDDM prevalences for some populations of North America, where estimates of Amerindian admixture are available from allele frequency surveys. Figure 3 presents the relationship between the proportion of Amerindian genes in these populations and their NIDDM prevalences. It is clear that the association between m (the proportion of Amerindian ancestry in the population) and d_H (NIDDM prevalence) is too conspicuous to be explained by chance alone (Kendall's $\tau = 0.848 \pm 0.221$, $P = 8.1 \times 10^{-5}$), suggesting that there is a strong correspondence in NIDDM prevalence with the extent of Amerindian ancestry.

Some comments about this compilation are in order before formally fitting any of the models to these data. First, not all of these surveys estimated the disease prevalence and the degree of Amerindian ancestry from the same set of individuals. Second, the source populations from which these admixed

TABLE 3. Prevalence of NIDDM and extent of Amerindian ancestry in some populations of North America

Population	No. surveyed	No. affected	Prevalence[1]	Percentage of Amerindian ancestry[1]
1. Pima (Arizona)	1,210	605	0.500 (1)	99.10 (2)
2. Seminoles (Florida)	254	89	0.350 (3)	91.87 (4)
3. Seminoles (Oklahoma)	217	84	0.387 (3)	85.93 (4)
4. Cherokee (North Carolina)	448	130	0.290 (5)	63.00 (6)
Mexican-Americans of San Antonio Males				
5. Barrio	178	27	0.152 (7)	45.20 (8)
6. Transitory	189	25	0.132 (7)	26.13 (8)
7. Suburb	184	12	0.065 (7)	23.00 (8)
Mexican-Americans of San Antonio Females				
8. Barrio	298	49	0.164 (7)	49.61 (8)
9. Transitory	243	17	0.070 (7)	35.62 (8)
10. Suburb	196	6	0.031 (7)	21.81 (8)
11. Mexican-Americans (Starr County)	1,930	217	0.112 (9)	29.94 (10)
12. Non-Hispanic Caucasians (San Antonio)	929	55	0.059 (7)	0.00 (8)

[1]Sources are in parentheses: 1, Knowler et al. (1981); 2, Williams et al. (1985); 3, Elston et al. (1974); 4, Pollitzer et al. (1970); 5, Stein et al. (1965); 6, Pollitzer et al. (1962); 7, Stern et al. (1984); 8, Chakraborty et al. (1986); 9, Hanis et al. (1983); and 10, Hanis et al. (1986).

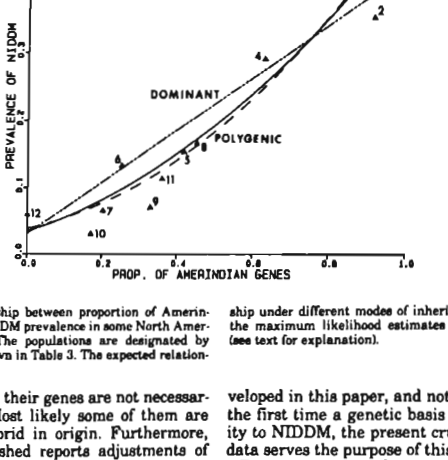

Fig. 3. Relationship between proportion of Amerindian genes and NIDDM prevalence in some North American populations. The populations are designated by serial numbers shown in Table 3. The expected relationship under different modes of inheritance are based on the maximum likelihood estimates of the parameters (see text for explanation).

groups received their genes are not necessarily identical. Most likely some of them are not truly dihybrid in origin. Furthermore, from the published reports adjustments of crude prevalence rates for the effects of age, sex, and obesity were not possible. Nevertheless, since the present analysis is primarily to illustrate the use of the methodology developed in this paper, and not to discover for the first time a genetic basis for susceptibility to NIDDM, the present crude form of the data serves the purpose of this investigation.

To see whether the correspondence between NIDDM prevalence and degree of Amerindian ancestry can be explained by any simple genetic model of susceptibility to

NIDDM, we fitted equations 17, 20, and 27 to the data presented in Table 3. Such model fitting required estimation of relevant parameters of the models (d_1 and d_2 for recessive and polygenic models, and d_1, d_2, and c for the dominant model). Estimation was carried out evaluating the total likelihood for the entire data, considering the 12 surveys as independent binomial samples and maximizing the total likelihood as a function of the parameters.

The total likelihood is given by

$$L = \prod_{i=1}^{12} \binom{N_i}{n_i} P_i^{n_i} (1 - P_i)^{N_i - n_i} \quad (28)$$

or

$$\ell n\, L = \sum_{i=1}^{12} \ell n \binom{N_i}{n_i} + \sum_{i=1}^{12} [n_i\, \ell n\, P_i + (N_i - n_i)\, \ell n\, (1 - P_i)], \quad (29)$$

where N_i and n_i are the total number of individuals surveyed and the number of individuals affected in the i-th survey, and

$$P_i = \begin{cases} m_i^2 d_1 + (1 - m_i)^2 d_2 + 2m_i(1 - m_i)\sqrt{d_1 d_2} \\ \quad \text{for the recessive model,} \\ m_i^2 d_1 + (1 - m_i)^2 d_2 + 2m_i(1 - m_i) \times \\ [(1 + c) - \sqrt{(1 + c - d_1)(1 + c - d_2)}] \\ \quad \text{for the dominant model,} \\ 1 - \Phi[m_i \Phi^{-1}(1 - d_1) + (1 - m_i)\Phi^{-1}(1 - d_2)] \\ \quad \text{for the polygenic model.} \end{cases} \quad (30)$$

m_i being the proportion of Amerindian genes in the i-th population.

Maximization of ℓn L (equation 29) was conducted by grid search method in the range $0 \leq d_1 < d_2 \leq 1$, and $c \geq 0$. The maximum likelihood estimates of d_1 and d_2 for the recessive and polygenic models are 0.481, 0.034, and 0.488, 0.038, respectively, yielding ℓn L values of -2746.95 and -2742.98, ignoring the constant term

$$\sum_{i=1}^{12} \ell n \binom{N_i}{n_i}. \quad (31)$$

The dominant model yielded the maximum likelihood estimates $\hat{d}_1 = 0.417$, $\hat{d}_2 = 0.031$, and $\hat{c} = 10.0$, with ℓn L $= -2787.95$. The unconstrained likelihood of the data (using $\hat{P}_i = n_i/N_i$) resulted in ℓn L $= -2720.82$. Since none of the models are submodels of one another, instead of the traditional likelihood ratio test (to examine the relative performance of the models), we used a goodness-of-fit test criterion. For each population survey, therefore, we used the estimated parameter values ($\hat{d}_1$, $\hat{d}_2$, and $\hat{c}$) to predict the probability of affection (NIDDM) for the given m_i values, and computed the expected number of affected ($\hat{n}_i = N_i \times \hat{P}_i$). Table 4 presents these estimates and the goodness-of-fit χ^2 values, where

$$\chi^2 = \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}. \quad (32)$$

It is clear that for each of the models the fit for the pooled data did not reach statistical adequacy, since the expected number of affecteds differed from the observed significantly ($P < 0.05$) in at least three of the 12 surveys in each of the models examined. The total χ^2 values (with 12 df) were also significant ($P < 0.01$) for all the three models.

In Figure 3 we plotted the expected relationship between m and d_H, assuming that each of these populations are derived from Caucasian and Amerindian gene pools and using the parameter estimates as noted above. The admixture estimates presented in Figure 3 are taken from the literature (sources indicated in the footnote of Table 3). These estimates are all based on allele frequency surveys. Both the prevalence figures and the admixture estimates are subject to sampling errors (e.g., Non-Hispanic Caucasians have a prevalence rate between 2 and 10%, depending on age), but it is interesting to note that the fits of the recessive and polygenic models are notably close, while the dominance curve seems to overestimate the disease prevalence for small values of m. All of the data points are within the 95% confidence bounds of the predictions of these genetic models for the sample sizes from which the disease prevalence values are computed. In terms of the goodness-of-fit test criterion, however, we have to choose between a recessive or a polygenic model as the best fit to the data, as shown in Table 4. In view of a smaller number of parameters in the recessive model, this might be viewed as the parsimonious model, since the fit is not statistically improved by fitting a more involved polygenic model.

TABLE 4. Predicted NIDDM prevalence in the different populations under different modes of inheritance and the goodness-of-fit χ^2 test criteria

Population	Observed No. of affected	Expected No. (E) and χ^2					
		Dominant		Recessive		Polygenic	
		E	χ^2	E	χ^2	E	χ^2
1. Pima (AZ)	605	500.0	22.0**	574.3	1.6	582.9	0.8
2. Seminol (FL)	89	97.9	0.8	108.0	3.4	109.7	3.9*
3. Seminol (OK)	84	78.7	0.4	83.9	0.0	85.0	0.01
4. Cherokee (NC)	130	123.0	0.3	114.3	2.2	111.9	2.9
Mex-Amer. Males (TX)							
5. Barrio	27	36.7	2.6	30.6	0.4	28.9	0.1
6. Transitional	25	25.0	0.0	19.0	1.9	17.7	3.0
7. Suburb	12	22.1	4.6*	16.7	1.3	15.6	0.8
Mex-Amer. Females (TX)							
8. Barrio	49	66.5	4.6*	56.9	1.1	54.1	0.5
9. Transitional	17	341.0	14.1**	32.5	7.4**	30.2	5.8*
10. Suburb	6	22.6	12.2**	17.1	7.2**	16.0	6.3*
Starr County (TX)							
11. Mex-Amer.	217	283.6	15.6**	219.0	0.02	203.2	0.9
12. Non-Hispanics	55	28.8	23.9**	31.8	17.0**	35.4	10.8**
Total (χ^2) with 12 df			101.1**		43.5**		35.9**

The parameter values used in computing the expected prevalence are dominant model, $d_1 = 0.417$, $d_2 = 0.031$, $c = 10.0$; recessive model, $d_1 = 0.481$, $d_2 = 0.034$; polygenic model, $d_1 = 0.488$, $d_2 = 0.038$.
*$P < .05$.
**$P < .01$.

From these data, we conclude that even though this relatively crude data compilation cannot conclusively demonstrate the mode of inheritance of NIDDM, there is a clear indication that genetic factors are largely responsible for susceptibility to NIDDM. The only caveat is that if, in these populations, the environmental factors responsible for NIDDM change in parallel with the extent of Amerindian ancestry, our inference may be in error. This possibility is very unlikely given the varied geographic and cultural origin of the data. In the San Antonio populations, we examined this in some detail (see Chakraborty et al., 1986), where we find that the major component of gene diversity in this population is due to diabetic status of individuals, not to socioeconomic classification. The genetic basis of NIDDM had been postulated through more incisive studies before (see, e.g., Elston et al., 1974; Beaty et al., 1982; Bennett, 1986).

DISCUSSION

The theory presented here demonstrates that there are distinct trends in the relationship between the proportion of accumulated admixture and disease frequency for disease with well-defined modes of inheritance. Even though this shows the importance of admixture studies designed to reveal the existence or nature of genetic factors in the etiology of a chronic disease, applications of the above formulations should be made with caution. Among the assumptions involved, the one we consider crucial is the assumption of constancy of the sporadic rate (in the case of simple mendelian inheritance) or if heritability (in the case of multifactorial threshold model) over all populations. While this might be a reasonable assumption for some cancers (see Vogel, 1979) and for some quantitative physiological traits (see Sing and Orr, 1978; Cloninger et al., 1983), this assumption implies that the relevant environmental factors are uniform over the entire array of populations studied. This is almost surely not the case for the populations examined in Table 3 and Figure 3. The fact that NIDDM prevalence does not depend solely on the extent of Amerindian ancestry is evident from the fact that its prevalence varies greatly among different Amerindian tribes and has probably changed (risen) dramatically in the past few decades (Bennett, 1983; Weiss et al., 1984a,b).

Our Figure 3 is consistent with the hypothesis that the relationship between Amerindian admixture and NIDDM prevalence rate may be explained by a simple model such as a major effect of a single recessive locus with some degree of sporadic occurrences. However, the fact that earlier prevalence studies did not reveal a high occurrence of NIDDM in pre-World War II populations of Amerindians, together with the evidence that several Amerindian groups are beginning to show increased rates of diabetes with "westernization" of their diet, habit, and customs clearly shows that the concept of gene–environment interaction must be quite important in the etiology of NIDDM (Bennett, 1983; Weiss et al., 1984a,b). Nevertheless, the above conclusion is in accordance with the complex segregation analysis of Elston et al. (1974) and Beaty et al. (1982). In both of these studies, pedigree analysis suggested the predominant effect of a single major locus. While Elston et al. (1974) failed to reject a dominant model in Oklahoma Seminole Indians, they preferred a recessive major locus model from distribution considerations of glucose tolerance values. Beaty et al. (1982), on the other hand, suggested a dominant single locus model, with perhaps some degree of common sibship–shared environment. Bennett (1986) reported segregation analysis results in Pima Indians of Arizona consistent with a single dominant locus mode of transmission with perhaps some effects of gene dosage associated with minor polygenic components.

The genetic basis of NIDDM has also been discussed from the standpoint that the degree of Amerindian ancestry in admixed groups of North America may be part of a "syndromic" appearance of a group of physiologically related disorders (obesity, diabetes, gallbladder disease, retinopathy, and so forth), which has been termed the "New World Syndrome" (Weiss et al., 1984a,b). The present study of association of NIDDM prevalence with the amount of Amerindian admixture supports the view that diabetes as a part of this larger array of metabolic disorders is most likely being expressed in individuals with Amerindian ancestry exposed to "western" environment.

One might argue that the extent of exposure to western habits of diet and custom may be correlated with, or parallel, the decrease of Amerindian genes in an admixed group, such that the relationship we observe may be explained by an environmental model only. Yet decreasing admixture is associated with increased westernization and yet de-

creased disease prevalence. Furthermore, there would have to be a fine-tuning of environmental changes with regard to the degree of admixture, such that it followed a well-specified rule like the genetic models examined here. This seems unlikely. This is particularly so, since customary habits and diet may change several times during the process of cultural confrontation with "western" people, while the change of accumulated admixture by the process of exchange of genes between gene pools is far slower.

We are not suggesting here that one can statistically "fit" a genetic model to data on differential population admixture, because there are other variables, usually uncontrolled in the samples, that may effect the relationship between admixture and disease occurrence. However, if a specific model seems to fit the data well, this may indicate the plausibility of a genetic explanation and suggest that the genetic hypothesis is worthy of more detailed and specific study. An initial study design aimed at finding regular admixture–disease prevalence patterns would be a useful first step.

This study also shows some of the problems in interpreting data of this sort. Among other things, the prevalance of NIDDM in Amerindian populations is rapidly rising, so that many of the data points are, or will be, out of date, whereas the theory is essentially one of a steady state. As mentioned earlier, the data analyzed here do not address the issue of incomplete penetrance of susceptibility genes for diabetes, and the prevalence rates are not adjusted for concomitant variables like age, sex, and obesity. Furthermore, the genetic models used here are somewhat oversimplified ones. For example, we assumed that c, the sporadic versus genetic occurrence ratio, is the same in all populations. This is possibly not correct, since as gene frequency approaches 1.0 (e.g., in Amerindians), the genetic fraction approaches 1.0, and the ratio (c) must approach zero. In addition, at least in the case of NIDDM, there are circumstantial reasons to believe that the ratio of genetic versus environmental variances of liability may fluctuate considerably across populations.

The theory presented here is not intended to provide alternatives for family data analysis of diseases of complex etiology. In fact, in comparison to exploratory data analysis (SEDA) or complex segregation analysis (Karlin et al., 1982; Elston, 1980; Morton et al., 1983), the power of discrimination of different modes of inheritance by the present approach is much smaller. The theory discussed here, for the first time, rationalizes the approach of "migrant studies" in epidemiology looking for genetic basis of chronic diseases. It shows that the use of specific genetic models, and of data on disease occurrence in populations with differing levels of admixture from source populations of known disease prevalence, should be a useful screening strategy in genetic epidemiology.

ACKNOWLEDGMENTS

This work was supported by U.S. Public Health Service research grants GM 20293, AM 27582, and AM 34666 from the National Institutes of Health. Comments from Drs. W.J. Schull, C.L. Hanis, and P.E. Smouse and three anonymous reviewers greatly increased our understanding of the problem studied.

LITERATURE CITED

Beaty, TH, Neel, JV, and Fajans, SS (1982) Identifying risk factors for diabetes in first degree relatives of non-insulin dependent diabetic patients. Am. J. Epidemiol. *115*:380–397.

Bennett, PH (1983) Diabetes in developing countries and unusual populations. In: JI Mann, K Pyörälä, and A Teuscher (eds): Diabetes in Epidemiological Perspective. Edinburgh: Churchill Livingstone, pp. 43–57.

Bennett, PH (1986) Chronic diseases in the American Indian, with particular reference to the Pima Indians of Arizona. Paper presented at the Woodlands Conference on Chronic Diseases Among Mexican Americans, December 2–6, 1984 (in press).

Chakraborty, R (1986) Relative contributions of gene and environment: Attributable risk and heritability. J. Ind: Anthropol. Soc. (in press).

Chakraborty, R, Ferrell, RE, Stern, MP, Haffner, SM, Hazuda, HP, and Rosenthal, M (1986) Relationship of prevalence of non-insulin dependent diabetes mallitus with Amerindian admixture in the Mexican Americans of San Antonio, Texas. Genet. Epidemiol. (in press).

Chakraborty, R, and Szathmary, EJE (eds) (1985) Diseases of Complex Etiology in Small Populations: Ethnic Differences and Research Approaches. New York: Alan R. Liss, Inc.

Chakravarti, A, and Chakraborty, R (1978) Elevated frequency of Tay Sachs disease among Ashkenazic Jews unlikely by genetic drift alone. Am J. Hum. Genet. *30*:256–261.

Cloninger, CR, Rao, DC, Rice, J, Reich, T, and Morton, NE (1983) A defense of path analysis in genetic epidemiology. Am. J. Hum. Genet. *35*:733–756.

Crittenden, LB (1961) Interpretation of familial aggregation based on multiple genetic and environmental factors. Ann. N.Y. Acad. Sci. *91*:769–780.

Elston, RC (1980) Segregation analysis. In: JH Mielke and MH Crawford (eds): Current Developments in Anthropological Genetics. New York: Plenum Press, Vol. 1, pp. 327–354.

Elston, RC, Namboodiri, KK, Nino, HV, and Pollitzer, WS (1974) Studies on blood and urine glucose in Seminole Indians: Indications for segregation of a major gene. Am. J. Hum. Genet. *26*:13–34.

Falconer, DS (1960) Introduction to Quantitative Genetics. New York: Roland Press.

Falconer, DS (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann. Hum. genet. *29*:51–76.

Feller, W (1967) An Introduction to Probability Theory and Its Applications. New York: John Wiley and Sons, Vol. I.

Gardner, LI, Stern, MP, Haffner, SM, Gaskill SP, Hazuda, HP, Relethford, JH, and Eifler, CW (1984) Prevalence of diabetes in Mexican Americans: Relationship of percent of gene pool derived from native American sources. Diabetes *33*:86–92.

Haenszel, W (1982) Migrant studies. In D Schottenfeld and JF Fraumeni (eds): Cancer Epidemiology and Prevention. Philadelphia: W.B. Sunders Company, pp. 194–207.

Hanis, CL, Chakraborty, R, Ferrell, RE, and Schull, WJ (1986) Individual admixture estimates: Disease associations and individual risk of diabetes and gallbladder disease among Mexican Americans in Starr County, Texas. Am. J. Phys. Anthropol. (in press).

Hanis, CL, Ferrell, RE, Barton, SA, Aguilar, L, Garza-Iberra, A, Tulloch, BR, Garcia, CA, and Schull, WJ (1983) Diabetes among Mexican Americans in Starr County, Texas. Am. J. Epidemiol. *118*:659–672.

Karlin, S, Williams, PT, Barrett-Connor, E, Hoover, JJ, Fraquhar, JW, What, WP, Haskell, WL, Bergelin, R, and Swarez, L (1982) Methods for the study of familial associations with applications to height, weight, lipid, and lipoprotein measurements. Am. J. Epidemiol. *116*:1001–1021.

Kirk, RL, Serjeantson, SW, King, H, and Zimmet, P (1985) The genetic epidemiology of diabetes mellitus. In: R Chakraborty and EJE Szathmary (eds): Diseases of Complex Etiology in Small Populations: Ethnic Differences and Research Approaches. New York: Alan R. Liss, Inc., pp. 119–146.

Knowler, WC, Pettitt, DJ, Savage, PJ, and Bennett, PH (1981) Diabetes incidence in Pima Indians: Contributions of obesity and parental diabetes. Am. J. Epidemiol. *113*:144–156.

Lilienfeld, AM, and Lilienfeld DE (1982) Foundations of Epidemiology. New York: Oxford University Press.

Morton, NE, Rao, DC, and Lalouel, J-M (1983) Methods in Genetic Epidemiology. Basel: Karger.

Mourant, AE, Kopec, AC, and Domaniewska-Sobczak, K (1976) The Distribution of the Human Blood Groups and Other Polymorphisms, 2nd ed. London: Oxford University Press.

Pollitzer, WS, Hartman, RC, Moore, H, Rosenfield, RE, Smith, H, Hakim, S, Schmidt, PJ, and Leyshon, WC (1962) Blood types of Cherokee Indians. Am. J. Phys. Anthropol. *20*:33–43.

Pollitzer, WS, Namboodiri, KK, Elston, RC, Brown, WH, and Leyshon, WC (1970) The Seminole Indians of Oklahoma: Morphology and serology. Am. J. Phys. Anthropol. *33*:15–29.

Ramot, B (1974) Genetic Polymorphisms and Diseases in Man. New York: Academic Press.

Sing, CF, and Orr, JD (1978) Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. IV. Separation of polygene from common environmental effects. Am. J. Hum. Genet. *30*:491–504.

Stein, J, West, K, Robey, J, Tirador, D, and McDonald, G (1965) The high prevalence of abnormal glucose tolerance in the Cherokee Indians of North Carolina. Arch. Intern. Med. *116*:842–845.

Steiner, PE (1954) Cancer: Race and Geography. Baltimore: Williams and Wilkens.

Stern, MP, Rosenthal, M, Haffner, SM, Hazuda, HP, and Franco, LJ (1984) Sex difference in the effects of sociocultural status on diabetes and cardiovascular risk factors in Mexican Americans: The San Antonio heart study. Am. J. Epidemiol. *120*:834–851.

Vogel, F (1979) Genetics of retinoblastoma. Hum. Genet. *52*:1–54.

Wagener, DK, and Cavalli-Sforza, LL (1975) Ethnic variation in genetic disease: Possible roles of hitchhiking and epistasis. Am. J. Hum. Genet. *27*:348–364.

Wagener, DK, Cavalli-Sforza, LL, and Barakat, R (1978) Ethnic variation of genetic disease: Roles of drift for recessive lethal genes. Am. J. Hum. Genet. *30*:262–270.

Weiss, KM, Ferrell, RE, and Hanis, CL (1984a) A new world syndrome of metabolic diseases with a genetic and evolutionary basis. Yearbook Phys. Anthropol. *27*:153–178.

Weiss, KM, Ferrell, RE, Hanis, CL, and Styne, PN (1984b) Genetics and epidemiology of gallbladder disease in New World native peoples. Am. J. Hum. Genet. *36*:1259–1278.

Williams, RC, Steinberg, AG, Gershowitz, H, Bennett, PH, Knowler, WC, Petitt, DJ, Butler, W, Baird, R, Dowda-Rea, L, Burch, TA, Morse, HG, and Smith, CG (1985) GM allotypes in native Americans: Evidence for three distinct migrations across the Bering land bridge. Am. J. Phys. Anthropol. *66*:1–20.

Zimmet, P (1982) Type 2 (Non-insulin-dependent) diabetes: An epidemiological overview. Diabetologia *22*:399–420.