THE PROBABILITY OF CONSANGUINEOUS MARRIAGES

L. L. CAVALLI-SFORZA, M. KIMURA[1], AND I. BARRAI

International Laboratory of Genetics and Biophysics, Pavia Section, Istituto di Genetica, University of Pavia, Pavia, Italy

Received December 22, 1965

THE frequency with which a husband is related to his wife, which we call the probability of consanguineous marriage, is determined by a number of factors, among which the following seem to be the most important ones: (1) the abundance of relatives, which depends on the type of relationship and on population growth. (2) The availability of consanguineous individuals in the "mating range", (migration causes a dispersal of relatives whose effect increases as the relationship becomes more remote: the more migration, the less consanguineous marriage). (3) The availability of the consanguineous individuals in the right age groups (age effect). (4) Assortive mating for socio-economic conditions and physical traits may have to be considered because of the similarity between relatives. (5) Traditions for or against some types of consanguineous marriages may also be a factor of importance. (6) There may exist other factors of social or economic nature.

In an earlier paper (BARRAI, CAVALLI-SFORZA and MORONI 1962), we showed the influence of factors 2 and 3. Another effect was also found, belonging to group 6. In the present paper we will concentrate on evaluating the effect of the first four factors, with a view to estimating the probability of consanguineous marriage in a population for which some necessary demographic parameters are available.

The necessary parameters are essentially those specifying the distributions of the distance between birth places, as well as of the age differences: between sibs, between father and offspring, between mother and offspring, between husband and wife. As estimates of these parameters are not usually available, a sample survey was carried out in an area (in the Parma province) for which information on consanguineous marriages was already at hand (MAINARDI, CAVALLI-SFORZA and BARRAI 1962). Data from the sample survey will be used in this paper.

The number of relatives: We will consider the simplest, and commonest type of relationship represented by individuals who are related via two sibs, as in Figure 1. In that example, the chains of descent via two sibs lead to the two relatives A and B, who are second cousins once removed. We will call i the number of ancestors between the common ancestors and the male relative A, j the number between the common ancestors and the female relative, B. Thus, in Figure 1, $i = 3$, $j = 2$, and $n = i + j$ is the number of *intermediate* ancestors. Of these, n_0

[1] On leave from the National Institute of Genetics, Mishima, Sizuoka-Ken, Japan. Present address: The same institute. This paper also constitutes Contribution No. 606 from the National Institute of Genetics. Aided in part by a Grant-in-Aid for Fundamental Scientific Research from the Ministry of Education in Japan.

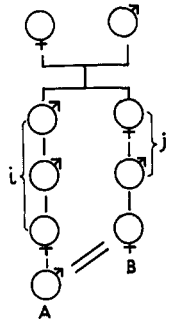


FIGURE 1.—Example of consanguineous mating: for definition of symbols, see text.

will be females and n_1 will be males, with $n_0 + n_1 = n$. In Figure 1, $n_0 = 2$, $n_1 = 3$.

If s is the expected number of sibs per individual, p the expected number of progeny per individual, and if there is no correlation in fertility, there are (ignoring sex) 2^2p^3s relatives of type A per B individual in the example illustrated, because B has 2^2 grandparents each of which has s sibs, each of which has p^3 great grandchildren. In general, there are 2^ip^js relatives of the wife in a consanguineous marriage which have the same consanguinity degree with her as her husband, and 2^jp^is relatives of the husband.

In a closed, stationary population in which everybody marries, the expected number of married progeny per couple is $p = 2$. In a stable population $p = 2a$, where a is the factor of increase per generation. The average number of sibs s depends on the distribution of progeny size. In fact, a family with progeny size p ascertained through the progeny is counted p times and each individual in the progeny has $(p-1)$ sibs. Then if $\phi(p)$ is the frequency of progeny size p, $p\,\phi(p)/\sum_{p=1}^{\infty} p\,\phi(p) \equiv \psi(p)$ is the frequency of $(p-1)$ sibs (FROTA-PESSOA 1957). The mean of the distribution given by ψ is equal to $\sum_{p=1}^{\infty}(p-1)\,\psi(p) = (V + \bar{p}^2 - \bar{p})/\bar{p}$ where $\bar{p}$ and V are the mean and the variance of the number of progeny. When this is distributed as in Poisson, the mean number of sibs is equal to the mean number of progeny $\bar{p}$. For distributions with variances higher than the mean, the mean number of sibs is higher than $\bar{p}$ (MAINARDI *et al.* 1962).

In many populations, however, it is so closely $s = p = 2$ that, if the variance and correlation for fertility can be ignored, the expected number of relatives n_c is 2^n for even cousins, namely, for cousins having $i = j$, and 2^{n+1} for cousins once removed $(i = j \pm 1)$. But these are also the numbers of pedigrees which can be distinguished on the basis of the sex of intermediate ancestors (BARRAI *et al.* 1962) and therefore each individual has one expected relative for each pedigree type ignoring the sex of the relative. Ignoring the type of pedigree, each individual has four first cousins, four relatives with $F = 1/8$ (uncles, aunts, nieces, and nephews), 16 first cousins once removed, 16 second cousins, 64 second cousins once removed, and 64 third cousins. These classes of relatives will be here called *degrees* of relationship; while the relationship specified by a type of pedigree as determined

by the arrangement of males and females among common ancestors will be called *type of relationship.*

Migration effect: 1. *The discontinuous case.* The study of migration demands a choice of the model of the geographic distribution of population. A first choice is that between continuous and discontinuous models. If we prefer a discontinuous distribution, for computing the probability of consanguineous marriage, we must have two matrices, X and M, both of order k, where k is the number of groups of people (villages, tribes, castes, etc) into which the population is clustered. Elements x_{ij} of matrix X specify the probability that for a given type of relationship a relative of an individual born in village i is born in village j. Row elements must add to unity in each row. In matrix M, elements m_{ij} specify the frequencies that of all marriages in the area, one member is born in village i and the other is born in village j. This matrix is symmetric because we ignore sex, and the sum of its elements is unity. The probability of consanguineous marriages corresponding to a given type of relationship, will then be

$$P = n_c \sum \frac{x_{ij}\, m_{ij}}{n_i} \tag{1}$$

where the sum is extended to all combination of i and j, n_i is the size of group i and n_c is the number of relatives of that type.

Matrix X is not easy to obtain from field data. It can be computed however, as a product of other matrices S, A_0, A_1, defined below, each representing one step in the path connecting one consanguineous individual to his consanguineous mate in the pedigree, on the assumption that migration in successive steps is independent and therefore can be treated as a Markov process. In fact, a correlation between migration steps may exist, especially because of stratification in socio-economic conditions, not accounted for by the grouping method employed, when this is for instance, a purely geographic one. We shall give the treatment for equal group size n, in a stationary population at equilibrium for migration, and discuss a possible generalization later.

The computation of matrix X from matrices S, A_0, A_1 will be shown for simplicity using an example, namely, the pedigree of second cousins once removed shown in Figure 1. A_1 is a matrix of the transition probabilities for father-offspring migration. Its element a_{ij} is the probability that a child of a father born in village i is born in village j, with row elements adding to unity in each row. Because of migration equilibrium and equal group size, also columns will add to 1. The A_0 matrix is the same transition matrix for mother-offspring. Matrix S is a transition matrix for sib migration, whose element s_{ij} is the probability that the sib of an individual born in village i is born in village j. Matrix S is symmetric with rows and columns adding to 1.

Then, the matrix X for the example of Figure 1 can be equated to the product

$$A'_0\, A'_1\, A'_1\, S\, A_0\, A_1 \tag{2}$$

where A_0' is the transpose of A_0, etc. The above matrix product is obtained by

following the path from A to B in Figure 1, via intermediate ancestors. Following the reverse path from B to A one obtains

$$A'_1 A'_0 S A_1 A_1 A_0 \tag{3}$$

which, by a well known theorem of matrix algebra, is shown to be the transpose of product (2). Note that S is symmetric.

The probability of consanguineous marriage P_c will then become

$$\frac{1}{N} \Sigma x_{ij} m_{ij} \tag{1'}$$

where N is the size of the individual group, and n_c is put equal to 1.

Since the m_{ij}'s are the elements of a symmetric matrix, it is immaterial if we use expression (2) or (3) for computing matrix X whose elements appear in (1').

The above treatment is based on the assumption of equal group size. On the other hand, the group size may be different in the actual case, but we can still apply the above theory by considering actual groups as collections of subgroups with approximately equal size.

Although matrices A_0, A_1, S are not difficult to obtain, they are not usually available. It may therefore be convenient sometime, as a first rough approximation, to use the method suggested by BARRAI *et al.* (1962) which is essentially the same as that followed by HAJNAL (1963), of ignoring the possibility of marriage in the group outside the one in which the individual is born, as this probability is often small, and use an average of the probabilities that an individual will have a child born in the same group, as the one in which he was born. We shall see later to what formulas this method leads, and their shortcomings.

The probability for a given degree of relationship should be obtained by adding up the probabilities for the various types belonging to this degree, as these differ one from the other. It is only if there is no difference between male and female migration rates that the expected frequency of consanguinity types that form them is independent of the proportion of the sexes among intermediate ancestors.

2. *The continuous case.* The use of a discontinuous model may be unsatisfactory if the population distribution is nearer to the continuous one. Also, much detailed knowledge is necessary if we want to use the discontinuous model. An approximation by a continuous model may therefore be useful, as it requires somewhat less detailed statistical information.

In analogy to the study of isolation by distance, put forward by SEWALL WRIGHT (1963), we might consider two types of continuous population distributions, a one-dimensional and a two-dimensional type. It may be noted, however, that the first type is far less frequently encountered, at least in a pure form, in human populations, and represents in any case a simpler model than the two-dimensional type. We have therefore concentrated our attention on a model of two-dimensional isotropic migration. The one-dimensional case could be obtained fairly easily as a simplified treatment, following the lines that we will give here for the two dimensional model. For isotropic migration in a two-dimensional

habitat given by coordinates x, y, the density function that the marriage of an individual born at the origin takes place with a mate born in (x, y) will be

$$f(x, y) = \frac{f(r)}{2\pi r} \tag{4}$$

where $r = (x^2 + y^2)^{1/2}$, $f(r)$ is the probability density that individual A marries an individual born at distance r from A's birthplace, and $-\infty < x, y < +\infty$.

Suppose that an individual, say a male A, is at the origin, and consider a small area dS $(= dx \cdot dy)$ around a point (x, y). The number of females in that area is $(D/2)dS$, where D is the population density, and $D/2$ the population density of females.

If we know a function $M_c(x, y)$ giving the probability density that one relative of A with a given type of relationship is born at point (x, y) and if there are expected to exist altogether n_c relatives of that type, the expected number of A's female relatives living in area dS will be

$$\frac{n_c}{2} M_c(x, y) dS \tag{5}$$

The probability of marriage between two individuals with given type of relationship will be

$$P = \iint_{-\infty}^{+\infty} \frac{n_c M_c(x, y)}{D} \frac{f(r)}{2\pi r} dx\, dy \tag{6}$$

Noting that $dxdy = rdrd\theta$ and integrating over all values of θ between 0 and 2π we have

$$P = \frac{n_c}{D} \int_0^{\infty} M_c(x, y)\, f(r) dr \tag{7}$$

The function $M_c(x, y)$ measures the dispersal of relatives and is the convolution of the following distributions: (1) The probability distribution of the distance between the birth places of the sibs which start the chains of relationship; (2) n probability distributions, each representing one generation in the chain of relationship starting with the two sibs, where $n = i + j$ is the number of intermediate ancestors, i.e.; the ancestors between the common ancestors and the consanguineous mates.

Since it is important to distinguish male and female migration, the n one-generation steps of migration will have to be subdivided into n_0 female, and n_1 male generations $(n_0 + n_1 = n)$.

In taking M_c as the convolution of $n + 1$ distributions it is assumed that there is no correlation between migration and successive generations and, in the absence of information, this might be taken as a first approximation.

In order to give $M_c(x, y)$ one must know the elementary distributions of which M_c is made. The contribution of sib-sib migration was neglected, because it is very modest with respect to the other components. Migration distributions of interest to genetics have been recently analyzed for European populations

(SUTTER and TRAN NIGOC TOAN 1957; LUU-MAU-THANH and J. SUTTER 1963; CAVALLI-SFORZA 1958, 1963) and are extremely skew. Perhaps the best fit was obtained with gamma distributions which, when fitted with respect to r, had exponents close to -1, and could not therefore permit convolutions to be obtained over two dimensions.

Accordingly, it was tried to fit distribution functions that would lend themselves more easily to obtain the M_c function. Two such functions are the exponential distribution, and a two-dimensional normal distribution (with equal variances for x and y). When expressed with respect to r, such distributions are given in (8) and (9):

"exponential" $$m_E(r) = k\, e^{-kr} \tag{8}$$

"normal" $$m_N(r) = \frac{r}{V} e^{-r^2/2V} \tag{9}$$

Neither of these functions seems to fit adequately the observed data. It is not unreasonable, however (considering the variety of means of transportation employed), to use sums of two or more of such functions. Fitting these distributions to the Parma data by numerical maximum likelihood, it was found that the sum of two exponentials

$$m_E(r) = phe^{-hr} + (1-p)ke^{-kr} \tag{10}$$

or the sum of three normals:

$$m_N(r) = \frac{pr}{V_S} e^{-r^2/2V_S} + \frac{qr}{V_M} e^{-r^2/2V_M} + \frac{(1-p-q)r}{V_L} e^{-r^2/2V_L} \tag{11}$$

fit the data reasonably well (Table 1).

TABLE 1

Distributions of birth distances for father-offspring (F-O), mother-offspring (M-O), husband-wife (H-W) pairs. Observed frequencies are given from a sample of families living in 1958 in the Parma province

	F-O			M-O			H-W		
Distance (km)	obs	trinorm*	biexp*	obs	trinorm	biexp	obs	trinorm	biexp
0–1.56	340	339.7	338.4	293	289.2	291.7	133	132.1	132.2
1.56–4.06	11	11.0	11.1	18	17.8	18.1	10	9.92	10.0
4.06–7.81	8	7.3	5.7	21	18.5	13.9	6	7.0	8.4
7.81–12.81	10	9.8	6.9	16	24.0	15.4	12	11.5	9.7
12.81–19.06	6	7.2	7.5	16	15.5	15.1	14	11.7	10.0
19.06–26.56	4	4.4	7.5	7	6.4	13.5	9	8.7	9.7
26.56–35.31	4	4.7	7.2	5	5.0	11.1	2	7.4	8.7
35.31–45.31	7	5.7	6.5	6	5.9	8.5	10	7.7	7.3
45.31–∞	24	24.1	23.1	21	20.7	15.7	19	19.0	19.0
Total	414			403			215		
$\chi^2_{[6]}$			5.6			12.63			8.93
$\chi^2_{[4]}$		0.70			3.15			5.28	

* Theoretical distributions are given by equations (10) for the biexponential and (11) for the trinormal, and the method of fitting is given in text. Parameters of the theoretical distributions are given in Table 2.

3. *"Sum of Exponentials" for the migration distributions.* Using distribution (10) for the elementary migration step, integral (6), giving the probability of consanguineous marriage under consideration of the sole migration effect, requires numerical integration. If we use the cartesian coordinate system (x,y) for isotropic migration in two dimensions, the density function for migration in one generation can be expressed as

$$m(x,y) = \frac{m_1(r)}{2\pi r} \quad \text{where } r = \sqrt{(x^2+y^2)} \tag{12}$$

If $C(u,v)$ is the characteristic function of $m(x,y)$ such that

$$C(u,v) = \iint_{-\infty}^{+\infty} e^{iux+ivy}\, m(x,y)\,dx\,dy \ , \quad (i=\sqrt{-1}), \tag{13}$$

we obtain, in terms of r,

$$C(u,v) = \int_0^{\infty} m_1(r) J_0(sr)\,dr \tag{14}$$

where J_0 is the Bessel function and

$$s = \sqrt{(u^2+v^2)} \tag{15}$$

In particular, when $m_1(r)$ is given by $m_E(r)$ of (10), the characteristic function reduces to

$$C(s) = \frac{ph}{(h^2+s^2)^{1/2}} + \frac{(1-p)k}{(k^2+s^2)^{1/2}} \tag{16}$$

If we consider the distribution of migration distances after n_0 female generations and n_1 male generations, with parameters p_0, h_0, k_0 and p_1, h_1, k_1 respectively, if the migrations in different generations are independent, the density function will be

$$\begin{aligned} M(x,y) &= \frac{1}{(2\pi)^2} \iint_{-\infty}^{+\infty} e^{-i(xu+yv)}\, C_1^{n_1}(s) C_0^{n_0}(s)\,du\,dv \\ &= \frac{1}{2\pi} \int_0^{\infty} C_1^{n_1}(s) C_0^{n_0}(s)\, s J_0(rs)\,ds \end{aligned} \tag{17}$$

where $C_0(s)$ and $C_1(s)$ are the characteristic functions of the distributions of the female and male migrations over one generation. Figure 2 illustrates some of the results of convolution based on (17).

Equation (17) gives us the M_c function desired for equation (6). We now have

$$\mathrm{P} = \frac{n_c}{2\pi D} \iint_0^{\infty} C_1^{n_1}(s) C_0^{n_0}(s)\, s J_0(rs) f(r)\,dr\,ds \tag{18}$$

If $f(r)$ is also given (see Table 1) as a sum of two exponentials with parameters p_m, h_m, k_m, then, noting that

$$\int_0^{\infty} e^{-h_m r} J_0(sr)\,dr = \frac{1}{(h_m^2+s^2)^{1/2}} \tag{19}$$

one obtains

$$\begin{aligned} P = \frac{n_c}{2\pi D} \Big\{ & \frac{p_m}{2} \int_0^{\infty} D_1^{n_1}(t) D_0^{n_0}(t)\, \frac{h_m\,dt}{(h_m^2+t)^{1/2}} + \frac{1-p_m}{2} \\ & \int_0^{\infty} D_1^{n_1}(t) D_0^{n_0}(t) \frac{k_m\,dt}{(k_m^2+t)^{1/2}} \Big\} \end{aligned} \tag{20}$$

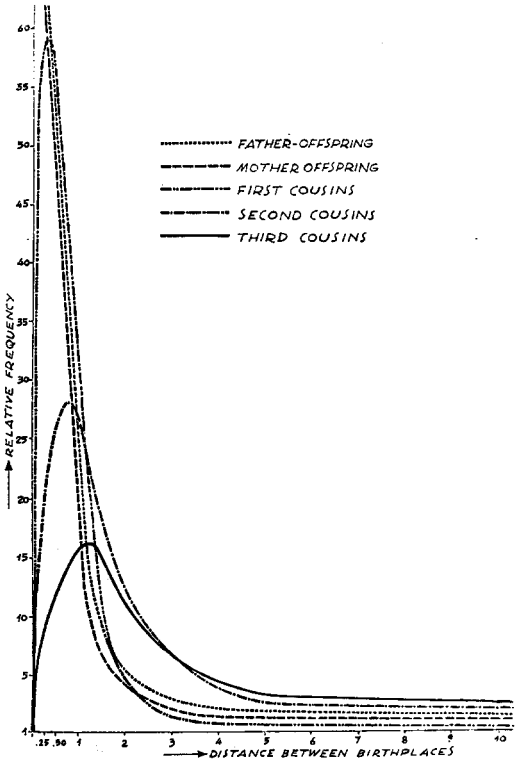


FIGURE 2.—Distributions based on equation (17).

in which

$$D_0(t) = \frac{p_0 h_0}{(h_0^2 + t)^{1/2}} + \frac{(1-p_0)k_0}{(k_0^2 + t)^{1/2}}$$
$$D_1(t) = \frac{p_1 h_1}{(h_1^2 + t)^{1/2}} + \frac{(1-p_1)k_1}{(k_1^2 + t)^{1/2}} \qquad (21)$$

Calling

$$I_m = P\,D \qquad (22)$$

and using for p, h, and k parameters the estimated set of values for male, female, and mating range distributions, expressions (21) can be used for the numerical evaluation of I_m, and hence, P.

4. *Sum of normals for fitting migration distributions:* If the distribution of migration distance is expressed by the sum of normal distributions, our calculation becomes much easier. This is mainly because the sum of any number of independent random variables, each of which is distributed normally, is again distributed normally with the mean and variance given respectively by the sums of means and variances of component normal distributions. Furthermore, if the distribution of migration in two dimensional cartesian coordinate (x,y) is given by normal distribution,

$$\frac{1}{2\pi V_i} e^{-(x^2+y^2)/2V_i} \tag{23}$$

and if the distribution of distance between mates is also given by another normal distribution,

$$\frac{1}{2\pi V_m} e^{-(x^2+y^2)/2V_m} \tag{24}$$

the integral of the product of the above two distributions can be expressed in the simple form:

$$I = \frac{1}{4\pi^2 V_i V_m} \iint_{-\infty}^{+\infty} e^{-(x^2+y^2)(V_i^{-1}+V_m^{-1})/2}\, dxdy = \frac{1}{2\pi(V_i+V_m)} \tag{25}$$

In what follows, we will assume that the individual migration is expressed by the sum of three normal distributions, corresponding to short range (S), medium range (M) and long range (L) components of migration. These distributions all have mean 0, but their variances are V_S, V_M and V_L respectively for short range, medium range and long range components. Thus the male migration in one generation may be expressed *symbolically*

$$p_1 S_1 + q_1 M_1 + (1-p_1-q_1) L_1 \tag{26}$$

where p and q are constants and S_1, M_1, and L_1 represent normal migration with variances, V_{S1}, V_{M1} and V_{L1} respectively. The corresponding expression for a female migration is

$$p_0 S_0 + q_0 M_0 + (1-p_0-q_0) L_0 \tag{27}$$

where S_0, M_0 and L_0 have variances V_{S0}, V_{M0} and V_{L0}. We will also assume that the distribution of distance between the birth places of mates in two-dimensional cartesian coordinate is expressed by the sum of three normal distributions:

$$f(x,y) = \frac{p_m}{2\pi V_{Sm}} e^{-(x^2+y^2)/2V_{Sm}} + \frac{q_m}{2\pi V_{Mm}} e^{-(x^2+y^2)/2V_{Mm}} + \frac{(1-p_m-q_m)}{2\pi V_{Lm}} e^{-(x^2+y^2)/2V_{Lm}} \tag{28}$$

This distribution will be expressed symbolically by

$$p_m S_m + q_m M_m + (1-p_m-q_m) L_m \tag{29}$$

With these expressions, the probability of marriage between two mates connected by n_1 generations of male migrations and n_0 generations of female migrations in their ancestors is

$$\mathrm{P} = \frac{n_c}{D} \iint_{-\infty}^{\infty} M_c(x,y) f(x,y)\, dxdy \tag{30}$$

in which $M_c(x,y)$ can be expressed symbolically by

$$M_c = [p_1 S_1 + q_1 M_1 + (1-p_1-q_1) L_1]^{n_1} [p_0 S_0 + q_0 M_0 + (1-p_0-q_0) L_0]^{n_0} \tag{31}$$

which is a sum of normal bivariate distributions:

$$M_c(x,y) = \sum_j c_j\, e^{-(x^2+y^2)/2V_j} / 2\pi V_j \ , \tag{32}$$

each term of this sum corresponding to a term of the expansion of (31). Each such term contains the symbolical product

$$S_1^a M_1^b L_1^{n_1-a-b} S_0^c M_0^d L_0^{n_0-c-d} \equiv T \ ; \tag{33}$$

which corresponds to a normal distribution whose variance is given by

$$V_j = aV_{S1} + bV_{M1} + (n_1-a-b)V_{L1} + cV_{S0} + dV_{M0} + (n_0-c-d)V_{L0} \quad (34)$$

while c_j in (32) will be given by the numerical coefficient accompanying the symbolical product (33) in the expansion. The integration in (30) may then be carried out using formula (25). Thus, the integral I_m in the right side of (30) may be expressed symbolically as follows:

$$I_m = M_c \cdot [p_m S_m + q_m M_m + (1-p_m-q_m) L_m] \quad (35)$$

and will give rise to a quantity

$$I_m = \sum_j b_j/(2\pi W_j) \quad (36)$$

which can be computed by expanding the symbolical product (35). The symbolical part of each term of the expansion contains $T \cdot Y$, where T is given by (33), and Y stands for either S_m, M_m or L_m. Corresponding to each TY there is a term in (36) with values $b_j = c_j\, p_m$ and $W_j = V_j + V_{Sm}$ if $Y = S_m$; $b_j = c_j\, q_m$ and $W_j = V_j + V_{Mm}$ if $Y = M_m$ and so on.

Age effects: Cousins of even degree are usually of similar age; cousins of uneven degree have usually some difference in age and therefore tend to marry less frequently. For an exact estimation of age effects, we need to know the probability distribution of age at marriage

$$d\,\phi_m(t,t') = \phi_m(t,t')\,dt\,dt' \quad (37)$$

where t' and t are the ages at marriage of male and female. We need also information on the distribution of the ages of consanguineous individuals.

$$\phi_c(t/t') \quad (38)$$

is the probability density of female (aged t) who are relatives of males aged t'. while

$$\phi_g(t) \quad (39)$$

is that of individuals from the general population, then

$$\psi_t = \phi_c/\phi_g \quad (40)$$

will express the frequency ratio of a certain age class among the relatives of an individual of age t' to the same age class in the general population.

This frequency ratio can be averaged over all the marriages taking place in the population by computing the integral

$$I_a = \iint \phi_m(t,t')\,\psi_t\,dtdt' \quad (41)$$

extended to the whole range of ages at marriage for t, and t'.

The function ϕ_g, namely the frequency of individuals of age t in the general population is given, to a first approximation, by a rectangular distribution which is constant over the range of ages of persons eligible for marriage. It may, however, decrease with increasing t in populations with high mortality in the reproductive period, or with decreasing birth rate, or it may also fluctuate as a consequence of irregularities of birth and death rates. If, however, population by ages were rectangular between 0 and ω,

$$\phi_g(t) = 1/\omega \quad (42)$$

where ω is the upper limit of the age distribution and

$$I_a = \omega \iint \phi_m(t,t')\,\phi_c(t,t')\,dtdt' = \omega J \tag{43}$$

where J represents the double integral.

The choice of males or females as the starting point in equation 38 is arbitrary and the procedure should be repeated after reversing the sexes, averaging the results. Since $\phi_c(t/t') = \phi_c(t'/t)$, it is enough to take as $1/\omega$ the average frequency per year of age, of individuals of either sex in the population, averaging over reproductive years.

In order to obtain the function (38) it is useful to obtain the distribution of age difference Δ between two relatives of different type of relationship. If there is no correlation between the age at marriage in various generations, the distribution of age difference Δ between relatives is easily computed. One needs, to this aim, information on: (1) the distribution of the age difference between sibs; as the order of birth of sibs is immaterial, this distribution has expected mean 0, and variance σ_s^2; (2) the distribution of generation times, namely the distribution of the age of the parent at birth of an offspring. These distributions are usually different for males and females and their means and variances will be given by the symbols τ_m, σ_m^2 for males and τ_f,σ_f^2 for females.

If we refer to Figure 1, we shall see that the computation of the expected age difference Δ between individual A and his relative B involves the difference between the sum of as many generation times τ as there are intermediate ancestors in the branch leading to B, minus the sum of as many generation times as there are intermediate ancestors in the branch leading to A. τ_m or τ_f, namely male or female generation times, must be taken each time depending on the sex of the intermediate ancestors. The age difference between sibs does not contribute to the expected value of Δ, because the order of birth of sibs is not taken into consideration, but it does contribute to the variance of Δ. Therefore, if m_i is the number of males among the common ancestors in the branch of the tree leading to A (the husband) and having i generations, and m_j is that in the other branch with j generations, (where $m_i + m_j = n_1$) the expected age difference is

$$\bar{\Delta} = E(\Delta) = (m_j - m_i)\tau_m + (j - i + m_i - m_j)\tau_f \tag{44}$$

and the variance of Δ will be given by

$$\sigma_\Delta^{\,2} = \sigma_s^2 + n_1\sigma_m^2 + (n - n_1)\sigma_f^2 \tag{45}$$

Formulas (44) and (45) were obtained by us (see BRAGLIA 1962) and also independently by HAJNAL (1963).

In order to provide material necessary for this type of evaluation, a sample of the population of the Parma province was subjected to analysis by questionnaire in 1958. The numerical results thus obtained will be used in the later part of this paper. Among other things, distributions of generation times and of the difference in age between sibs were derived. These distributions are somewhat asymmetric, but as Δ is the sum of several of them, it tends to normality rapidly with increasing $(i + j)$. A direct check of data available between first cousins showed good agreement with the expectation of normality (MAINARDI *et al.* 1966).

We can therefore use the following expression for the desired distribution given in (38) above:

$$\phi_c(t/t') = \frac{1}{\sigma_\Delta \sqrt{2\pi}} \exp[-(t'-t-\bar{\Delta})^2/2\sigma_\Delta{}^2] \tag{46}$$

If we want to proceed to obtain integral J given in (43) above, we need to specify the bivariate distribution of age at marriage, ϕ_m. This distribution is certainly not normal in most populations in which the data are available. We found it could be rather well represented by a normal correlated surface after transforming ages to $\log(t-t_{min})$ where t_{min} is the minimum age at marriage. However, the evaluation of j demanded in this case numerical integration.

When compared with the more exact treatment just mentioned, the results obtained by using the normal approximation to function ϕ_m were also satisfactory except for the uncle-niece or aunt-nephew case.

If we then consider ϕ_m to be a normal bivariate function with means μ_h, μ_w for the ages at marriage of husband and wife (where $\mu_h - \mu_w \doteq \tau_m - \tau_f$) respective variances σ^2_h, σ^2_w, and a correlation coefficient ρ_{hw}, then the integral J reduces to

$$J = \frac{1}{S\sqrt{2\pi}} e^{-M^2/2S^2} \tag{47}$$

where

$$M = \bar{\Delta} - (\mu_h - \mu_w) \doteq (m_j - m_i - 1)\tau_m + (j - i + m_i - m_j + 1)\tau_f \tag{48}$$

$$S^2 = \sigma_\Delta{}^2 + \sigma_h{}^2 - 2\rho_{hw}\sigma_h\sigma_w + \sigma_w{}^2 = \sigma_0{}^2 + n_1\sigma_m{}^2 + (n-n_1)\sigma_f{}^2$$

where

$$\sigma_0{}^2 = \sigma_s{}^2 + \sigma_h{}^2 - 2\rho\sigma_h\sigma_w + \sigma_w{}^2$$

a result comparable to that obtained by HAJNAL (1963).

It is interesting to compare the results given in (47) with observations in the paper by BARRAI *et al.* (1962) on the problem of age effects. The analysis summarized in Figures 1 and 2 of that paper indicated an approximately parabolic relationship between the logarithm of the frequency of a given type of consanguineous marriage and a function indicated in the abscissa which is a linear transformation of the quantity M given in (48) above. It will be noted that formula (47) gives an exactly parabolic relationship between $\log J$ and M if there is no difference between the variance of male and female generation times ($\sigma_m{}^2$ and $\sigma_f{}^2$). It is very probable, therefore, that the deviation from a parabola observed in Figures 1 and 2 of BARRAI *et al.* (1962) is due to the fact that the variances of male and female generation times are unequal.

The computation of the quantity I_a from formula (43) thus supplies a correction factor for the expected frequency of a given pedigree type of consanguineous marriage which can be applied to the expected frequency computed on the basis of migration alone, provided that between migration and age there is no important correlation. The corrected probability will be:

$$P_c = P\,I_a \tag{49}$$

Assortative mating for heritable traits: Assortative mating for socio-economic conditions or other traits which are inherited via biological or social mechanisms may also affect the frequency of consanguineous matings because of the higher resemblance between relatives. In order to assess its influence, one needs knowledge of the correlation between husband and wife (ρ) and that between relatives (r_c) for the trait or traits responsible for assortative mating. We will assume that the trait is measured in such a scale that x, y, the male and female values respectively, are normal with mean 0 and standard deviation σ in the general population. The expected value y_c of the trait, in female relatives of individuals of trait x, will then be $y_c = r_c x$, with variance $\sigma^2(1-r^2_c)$, and therefore the frequency of relatives with trait y_c of individuals with trait x will be:

$$\phi_c(y_c|x) = \frac{\exp\left[-(y_c - r_c x)^2/2\sigma^2(1-r_c^2)\right]}{\sigma\sqrt{2\pi(1-r_c^2)}} \tag{50}$$

while the frequency of individuals with trait y in the general population will be

$$\phi(y) = \frac{e^{-y^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \tag{51}$$

The ratio between the frequency of individuals with trait y among the relatives of an individual with trait x, and the same frequency among individuals from the general population will be

$$\psi(y|x) = \phi_c(y|x)/\phi(y) \tag{52}$$

If $\phi\ (x, y)$ is the bivariate correlated distribution of the trait, with correlation ρ between husband and wife

$$\phi(x,y) = \frac{\exp[-(x^2 - 2\rho xy + y^2)/2(1-\rho^2)]}{2\pi\sigma^2\sqrt{1-\rho^2}} \tag{53}$$

and

$$\phi(x,y) = \phi(x)\,\phi(y|x) \tag{54}$$

the integral

$$I_s = \iint_{-\infty}^{+\infty} \phi(x,y)\psi(y|x)\,dx\,dy \tag{55}$$

will give the average over all marriages of the ratio ψ. On integration this is found to be

$$I_s = \frac{1}{1-\rho r_c} \tag{56}$$

Factor I_s can be used to multiply the probability of consanguineous marriages computed on the basis of migration and age (if the trait in question is independent of both migration and age), in order to correct the probability for the effect of assortative mating.

It is likely that the postulated independence does not exist for socio-economic conditions and migration, so that this method of evaluation may be valid only as a first approximation.

In any case, it would seem, from what little knowledge is available that the correction factor is not likely to be large. From data collected in the Parma region ρ is of the order of .5, r_c is not known, but must be low. r_c could be estimated from knowledge of ρ and of the correlation between parent and offspring for the trait.

If r_{PO}, r_{MO} and r_s are the correlation coefficients between father and offspring, mother and offspring, sib and sib respectively, the correlation between relatives with n_0 and n_1 female or male intermediate ancestors will be

$$r_c = r_s \, r_{PO}^{n_1} \, r_{MO}^{n_0} \tag{57}$$

Even assuming that $\rho = r_{PO} = r_{MO} = 0.8$ the correlation between first cousins would be $r_c = .18$, that between second cousins $r_c = .116$, leading to correction factors of 1.20 and 1.11.

For the traits determined entirely by additive genes $r_{PO} = r_{MO} = r_s = 0.5$, and if $\rho = .25$ the correction factor would be 1.03. It would take a great many independent heritable traits to make the correction factor important. It therefore seems likely that one can neglect assortative mating effects at the first approximation.

Agreement between theory and observation: A complete test of the theory just given would require demographic knowledge which is not available in the literature. Material which has been collected in the Parma province contains such information but the work of analysis is not complete, especially for the part regarding demographic data of the past two centuries (BARRAI, CAVALLI-SFORZA and MORONI 1964). Changes of demographic patterns are known to have taken place especially in the last and the present century, and the study of consanguineous marriages requires demographic knowledge valid for ancestral generations. Therefore, until data for earlier times than now available is at hand, no satisfactory test will be possible.

At the moment, the source of data coming nearest to the requirements is that from the Upper Parma River Valley. Here a sample of almost 500 families coming from various villages of the area was investigated by questionnaire (MAINARDI *et al.* 1962). Some of this material is already published (CAVALLI-SFORZA 1963; CAVALLI-SFORZA *et al.* 1964). This is, at the moment, the only source of information on distribution of distances between birth places, and age differences, in which we are interested, but it comes from a contemporary population living in the same area in which we have collected consanguineous marriages.

1. *Migration.* The migration data utilized in Table 1 are from the source just cited. Biexponential and trinormal distributions were fitted by a fully numerical version of maximum likelihood estimation, computing the likelihood of a set of trial values of the parameters, then obtaning first derivatives by recomputing likelihoods with small increments added to each parameter in turn. From this the information matrix could be calculated and corrections of the trial values obtained. The procedure was iterated until the increase in likelihood was negligible. In some cases there were difficulties in obtaining a good fit using maximum likelihood and another method was employed in which chi-square is minimized as follows. One defines plausible intervals for each parameter and on the basis of chi-square decides which half-interval to use as trial estimate. One then proceeds by taking the better half-interval for each parameter in a new cycle of computation, progressing in this way until the required precision is reached.

Maximum likelihood estimates for biexponential and for trinormal migration

TABLE 2

Parameters of the fitted distribution of Table 1

	F-O	M-O	H-W
Trinormal			
p	.839	.740	.646
q	.061	.163	.144
$1-p-q$	.100	.097	.210
V_S	.32	.36	.42
V_M	60.0	58.7	89.9
V_L	1890.	1608.	1182.
Biexponential			
p	.828	.725	.614
h	2.493	2.330	2.266
k	.0248	.0425	.0325

curves to the distribution of birth place distances of the pairs father-offspring, mother-offspring, and husband-wife are given in Table 2.

The distribution of distances between birthplaces of sibs showed such a high concentration in the zero class that this distribution was neglected throughout. This has the only disadvantage that, when using biexponentials, the integral of equation (20) does not converge for uncle-niece or aunt-nephew, while it would if the convolution had included the sib-sib migration. Therefore we are unable to give expected values for this class of relatives under the hypothesis of biexponential migration.

Table 3 shows the probabilities of consanguineous marriage, P computed from equation (30) for trinormal migration and from equation (20) for biexponential migration. The numerical integration necessary in the latter case was carried out by computer. The migration parameters are those given in Table 2 and the population density employed in the calculation is that valid on average for the Upper Parma River Valley, years 1860–1962, and is $D = 45.7$ inhabitants per square kilometer. All probabilities are given for $n_c = 1$. All consanguineous marriages from uncle-niece or aunt-nephew to third cousin are considered, but only pedigrees that have different probabilities are distinguished, namely those pedigrees in which the number of males n_1 among intermediate ancestors (n) varies from 0 to n.

It will be noted that there is a marked discrepancy between the trinormal and the biexponential, which is more serious the nearer the relationship. The cause of this is believed to be the difference in behavior of the two functions at the origin. In fact, even if the two functions are fitted to the same observed distributions, the class at the origin, which represents on the average some 75% of the observations, is fitted in the biexponential by a monotonically decreasing function which cuts the ordinate at a value different from 0, while for the trinormal the function used is zero at the origin and goes to a peak thereafter. It seems therefore that with data such as the present ones, in which there is an accumulation of most of the observations in the class at the origin, the choice of one or the other

TABLE 3

The probability of consanguineous marriage under assumption of a trinormal and of a biexponential migration distribution as given in Tables 1 and 2. Only migration effect is considered

Degree of relationship	n	n_1	P values	
			trinormal	biexponential
Uncle niece, aunt nephew	1	0	.002150	.....
		1	.002558	.....
First cousins	2	0	.001088	.003886
		1	.001278	.004542
		2	.001499	.005311
1½ cousins	3	0	.000613	.001635
		1	.000714	.001889
		2	.000830	.002182
		3	.000961	.002522
Second cousins	4	0	.000366	.000843
		1	.000425	.000966
		2	.000432	.001107
		3	.000570	.001270
		4	.000654	.001457
2½ cousins	5	0	.000227	.000490
		1	.000263	.000559
		2	.000304	.000628
		3	.000350	.000716
		4	.000404	.000817
		5	.000462	.000933
Third cousins	6	0	.000145	.000301
		1	.000168	.000342
		2	.000193	.000388
		3	.000222	.000441
		4	.000255	.000501
		5	.000294	.000570
		6	.000335	.000649

n is the number of intermediate ancestors and n_1 is the number of males among them.

continuous model is not an easy one. The choice cannot be done on the only evidence of the goodness of fit of the theoretical distribution. As the type of the distribution is critical in determining the expectations, it is believed that it will be preferable in cases like this to adopt, when adequate demographic data are available, the discontinuous model, using migration matrices which do not assume any specific form of migration distribution.

Other evidence shows, in any case, that the demographic data available from the present day populations show only qualitative agreement with the consanguinity data. When the values of probabilities of consanguineous marriages of Table 3 are plotted on a graph (Figure 3) it will be seen that P rises almost exponentially with the number of males for a given n. The slope of increase of log P is about log 1.14 from Figure 3 data. This slope should correspond to the logarithm of the quantity called c and estimated from actual data coming from

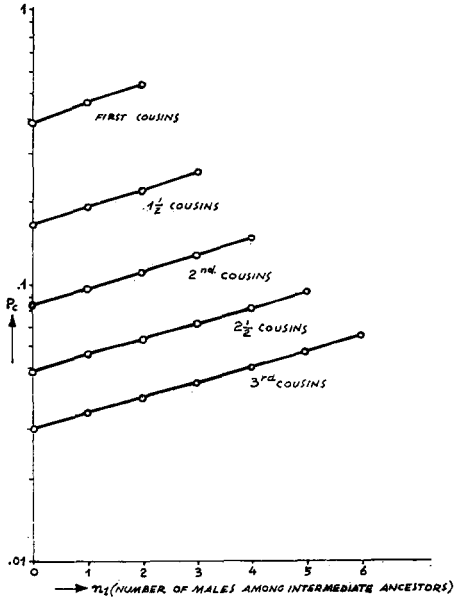


FIGURE 3.—Probabilities of consanguineous marriages, based on values in Table 3.

a comparable area in a former paper (BARRAI *et al.* 1962). There, values of c were, for the mountainous region of Parma, 1.389 for second cousins, 1.366 for third cousins and lower than 1 (.918) for first cousins, the last value being influenced (as shown in that paper) by a sociological factor which should account for the aberration. The agreement is thus only a qualitative one in the sense that the difference between male and female migration observed in a modern population, although in the right direction, is not as high as would be expected from the c value determined from second and third cousin marriage frequencies in the same area.

A priori the approach by the "sum of exponentials" is probably more satisfactory than that by the "sum of normals". In fact, the biexponential seems to represent a little more accurately the clustering of people in villages by its having the mode at the origin.

It is, however, of interest to follow further the "sum of normals" approach, because it does not require any numerical integration and is therefore more directly available for the calculation of expected frequencies of consanguinity. Although the migration integral I_m requires, with a trinormal distribution, the computation of 3^{n+1} terms, whose sum constitutes I_m, only one term is important in the present circumstances. This depends on the fact that the short range migration is so much more important in terms of frequency than the other two. It therefore happens that for the values given in Table 2, the term

$$\frac{p_1^{n_1} p_0^{n_0} p_m}{n_0 V_{s0} + n_1 V_{s1} + V_{sm}} \tag{58}$$

accounts for some 95% of the value of I_m. Since, approximately, $V_{s0} = V_{s1} =$

$V_{sm} = V_s$, formula (58) simplifies to

$$P \doteq \frac{1}{2\pi D} \frac{p_1^{n_1} p_0^{n_0} p_m}{V_s(n_1+n_0+1)} \tag{59}$$

It is interesting to compare this value with that obtained by the simplified discontinuous treatment, in which only marriages inside the village where both mates are born are considered. In the latter case let us take as P_1 the proportion of children born in the same village as their fathers, P_0 the proportion born in the same village as their mothers, P_m the proportion of husbands and wives who were born in the same village, and N the village size; then, always considering $n_c = 1$ and neglecting age and other corrections, for a pedigree with n_0 and n_1 female and male common ancestors,

$$P \doteq \frac{P_1^{n_1} P_0^{n_0} P_m}{N} \tag{60}$$

With the numerical values so far used (namely V_S approximately $= 0.4$, and $D = 45.7$) the two formulas from simplified normal migration and the simplified discontinuous treatment give approximately equal results only for first cousins, $(n = n_1 + n_0 = 2)$ as the average village size (N) in the corresponding region is not far from 300 and as $P_0 = .821$, $P_1 = .727$, $P_m = .645$, are not far from the corresponding values p_0, p_1, p_m of Table 2. For other types of relationships these two estimates inevitably diverge.

2. *Age effects:* A direct comparison between the P values given in Table 3 and the observed ones is not possible until we correct P values for age according to (49). To this aim, we need to know the required age distributions. From the 1958 Parma sample (MAINARDI *et al.* 1962) the values given in Table 4 were computed. Here also, however, we find a discrepancy between data from the contemporary population and those suited for the analysis of the consanguineous marriages, but again there is at least qualitative agreement. A test would be provided by the fitting of parabolas to the logarithms of the frequencies of consanguineous matings of various degrees versus the expected age difference between mates, as was done in Figures 1 and 2 of the paper by BARRAI *et al.* (1962), but a more direct test was used in Table 2 of the paper by CAVALLI *et al.* (1964), in which the demographic data given here in Table 4 were used to fit the frequencies of 29 classes of consaguineous matings from the Upper Parma Valley. Consanguineous marriages 1850 to 1950 were considered ranging from uncle-niece to second cousins, and grouped according to n_1, and n, which gives rise to the 29 classes shown in that table. Expectations of each class were computed as proportional to

$$P_m^{n_0} P_f^{n_1} e^{-M^2/2V} \tag{61}$$

where M, V are as defined by equations 48. When, however, the values of the contemporary sample, given in Table 4, were inserted in this equation, the fit to the observed consanguinity frequencies was not satisfactory. We therefore tried to fit P_m, P_f, τ_m, τ_f, σ^2_m, σ^2_f, $(\sigma^2_s + \sigma^2_\Delta)$ values to the data by numerical maximum likelihood. The expectations improved considerably and the only major dis-

TABLE 4

Estimates of parameters necessary for age corrections, obtained from a contemporary population in the Upper Parma River Valley (Mainardi et al. 1966)

Generation time:			
Males	mean	τ_m	33.24 ± 0.19
	variance	σ^2_m	46.60
Females	mean	τ_f	28.73 ± 0.16
	variance	σ^2_p	33.78
Variance of age difference between sibs		σ_s^2	25.78
Variance of age difference between mates		σ_o^2	33.31

crepancy left was between classes of first cousins, where other factors of sociological nature might be involved (CAVALLI-SFORZA *et al.* 1964). The demographic parameters thus estimated are given also here, in the heading of Table 6.

It will be noted that these estimates indicate a later average age at reproduction than in the modern population (by 2 or 3 years) as well as an increased variance of generation times. The consanguineous marriages from which these parameters were estimated extend from 1850 to 1950, and therefore, their intermediate ancestors lived mostly in the 19th Century or even earlier. The difference between the fitted values, which are estimated for earlier generations, and the values obtained for the present generation is probably in the right direction. Thus, removing geographical heterogeneity was not sufficient to give an agreement, other than qualitative, between our present estimates of the demographic parameters and those necessary for a good fit. It should be noted, in addition that in the period 1850 to 1950 some changes in the consanguineous frequencies were observed, and a more recent research (MORONI, in preparation) shows that very extensive changes in consanguineous frequencies took place in the 19th Century. Unfortunately, breaking down the figures further by periods of time would reduce them too much for a meaningful comparison with expected values to be possible. We shall have to be content at the moment with an approximate agreement, as is possible with the present data.

It may also be argued that the normal approximation to the distribution of ages at mating or at reproduction may be inadequate. However, it has been seen (MAINARDI *et al.* 1966) that for first cousins, age differences are normally distributed, in agreement with expectation, and the effect of the non-normality of the distribution of age at mating is noticeable only slightly for the extreme case of uncle-niece or aunt-nephew. It was found that ages at mating are well fitted by log $(t\text{-}t_{min})$ where t_{min} is the minimum legal age at marriage (HALD 1952). We have found that also the correlation surface between male and female ages at marriage is reasonably normal using the above transformation, and have used it to compute I_a, although it requires numerical integration. Table 5 shows that

TABLE 5

Age effect

Type of relationship	i	j	m_i	m_j	J values: Transformation log $(t\text{-}t_{min})$	J values: Untransformed ages	Difference %
Uncle-niece	0	1	0	0	.001688	.001649	2.4
	0	1	0	1	.000776	.000688	13.0
Aunt-nephew	1	0	0	0	.000146	.000133	10.0
	1	0	1	0	.000057	.000050	14.0
First cousin	1	1	0	0	.033564	.033006	1.7
	1	1	0	1	.034278	.034000	.8
	1	1	1	0	.026568	.026188	1.4
	1	1	1	1	.031437	.031030	1.3

Discrepancy between numerical integration using a normal bivariate distribution fitted to ages of mates transformed according to log $(t\text{-}t_{min})$ and direct integration using a normal bivariate distribution fitted to untransformed ages (1954 marriages, Italy)

the divergence between the two methods of computation is rather small, even in the most critical cases.

The quantity I_a from formula (43) gives the correction factor by which P values obtained from migration should be multiplied to obtain P_c. In order to obtain it, we need J from formula (47) and ω. The value ω will be computed here as the reciprocal of the mean frequency per time unit (years in this case) of individuals of either sex in the population, averaging over reproductive years. In practice, we have simply taken ages between 15 and 50 years and averaged the corresponding frequencies (unweighted for simplicity), using data from the 1901 Italian census, a time which corresponds to the middle of the period examined before. The value of ω thus obtained is 70.36 years.

Using as demographic values those given at the head of Table 6, the quantities $I_a = \omega J$ given in the body of Table 6 were obtained.

3. *Probabilities of consanguineous marriage.* It is now possible to obtain the P_c values corrected for age effects, multiplying each P_c value times the appropriate I_a value. The correspondence between P and I_a values is easily established, keeping in mind that m_i, m_j $(m_i + m_j = n_1)$ are the parameters specifying the number of males among intermediate ancestors in the branches leading respectively to the consanguineous husband and wife, and that i, j $(i+j=n)$ are the total numbers of intermediate ancestors in the two branches. An example of of calculations will be found in Table 7.

The result (always keeping $D = 45.7$ as in Table 3) in the calculation of P_c is given in Table 8, where both the biexponential and the trinormal models of migration are retained and compared with observations, after adding up for all pedigrees belonging to the same degree of relationship. When doing this, it should be remembered that some pedigrees are represented more than once, and that if $P_c(i,j,m_i,m_j)$ is the expected frequency of a pedigree with given i, j, m_i, and

TABLE 6

Correction factors for age (I_a) *computed from formulas (43), (47), and (48) using demographic values* estimated from frequencies of consanguineous marriages in the paper by Cavalli* et al. *(1964)*

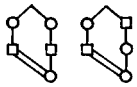
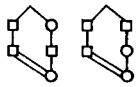
	m_i†	m_j	I_a
Uncle-niece	0	0	0.08851
	0	1	0.03018
Aunt-nephew	0	0	0.00316
	1	0	0.00091
First cousins	0	0	2.1688
	0	1	2.29962
	1	0	1.61077
	1	1	2.04407
First cousins once removed: husband older generation	0	0	0.33139
	0	1	0.16646
	0	2	0.07878
	1	0	0.66639
	1	1	0.38086
	1	2	0.20306
Second cousins once removed	0	0	0.14860
	0	1	0.29895
	0	2	0.52017
	1	0	0.07800
	1	1	0.17162
	1	2	0.32619
	2	0	0.03915
	2	1	0.09375
	2	2	0.19378
	3	0	0.01887
	3	1	0.04898
	3	2	0.10966
Third cousins	0	0	1.53060
	0	1	1.57073
	0	2	1.48415
	0	3	1.30132
	1	0	1.33034
	1	1	1.48415
	1	2	1.52047
	1	3	1.44166
	2	0	1.07619
	2	1	1.30132
	2	2	1.44166
	2	3	1.47475
	3	0	0.81582
	3	1	1.06610
	3	2	1.27388
	3	3	1.40262

* $\tau_m = 36.10$, $\tau_f = 30.95$, $\sigma^2_m = 53.32$, $\sigma^2_f = 42.59$, $\sigma^2_o = 53.08$, $\omega = 70.36$.

† As before, i,j are the numbers of intermediate ancestors in the branches of the pedigree leading to the husband and to the wife (see Figure 1) and m_i, m_j are the numbers of males among them.

TABLE 7

Computation of probabilities of consanguineous marriages P_c *in percent. Example for 1½ cousins, husband in shorter branch* (i = 1, j = 2)

Pedigree types	m_i	m_j	n'	I_m Trinormal	I_m Biexp	I_a	P_c percent Trinormal	P_c percent Biexp
	0	0	1	.000613	.001635	.33139	.0203	.0541
	0	1	2	.00714	.001889	.16646	.0119	.0314
	0	2	1	.000830	.002182	.07878	.0065	.0172
	1	0	1	.000714	.001889	.66639	.0476	.1259
	1	1	2	.000830	.002182	.38086	.0316	.0831
	1	2	1	.001961	.002522	.20306	.0195	.0512
						$\sum n'P_c =$	.1809	.4774

P data from Table 3 and I_a data from Table 6. m_i, m_j are the number of males among intermediate ancestors in the branches of the tree leading to the husband and wife respectively; n' is the number of pedigrees. P_c values are products $I_m \times I_a$ and are given in percent. In pedigrees, squares are males, circles females.

m_j, the sum of the frequencies of all pedigree types for a given pedigree $n = i + j$ will be given by:

$$P_c(n) = \sum^{i}(m_i)\,\sum^{j}(m_j) P_c\,(i, j, m_i, m_j) \qquad (62)$$

where the $\sum$ is extended to all pedigree types with different i, j, m_i, m_j values belonging to the same type of relationship. The expectations given as probabilities of consanguineous marriages in Table 8 are then obtained, corresponding to the two continuous models.

It will be noted that there is only an approximate agreement, but the migration estimates here employed are probably too large. This can account for the discrepancy becoming higher with less close relationship. In any case, the biexponential function gives a better fit. Also, the vagaries of demographic values and of consanguineous matings with time and geography are very probably responsible for a part of the discrepancy. As better estimates of the demographic parameters will be forthcoming in the not too distant future, we hope to have better opportunities for a more satisfactory test of the theory.

TABLE 8

Comparisons between probabilities of consanguineous marriages corrected for age (as in Table 7) and observed frequencies for upper Parma River Valley (communes of Corniglio, Monchio, Tizzano, Palanzano)

	m_i	m_j	Probability P_c		Observed marriages	
			Trinormal	Biexp	Number	Percent
Uncle-niece	0	0	.0190	....	5	.037
	0	1	.0077	....	4	.030
Aunt-nephew	0	0	.0007	....	1	.007
	0	1	.0002	....	0	0
Total			.0276			.074
First cousins	0	0	.2360	.8384	109	.8066
	0	1	.2939	1.0445	157	1.1618
	1	0	.2059	.7316	99	.7326
	1	1	.3052	1.0856	96	.7104
Total			1.0410	3.7001		3.4114
1½ cousins, in shorter branch	0	0	.1809	.4774	153	1.322
1½ cousins, in shorter branch	0	0	.0399	.0919	45	.3330
Second cousins	0	0	1.2870	2.8953	938	7.2520

It is also possible that some of the assumptions here made: independence of age and migration, lack of parent-offspring correlation in fertility, and in mobility, may limit the usefulness of the simple model here described, to an extent that further research may show.

This work has been supported by grants from the U.S. Atomic Energy Commission and by EURATOM-CNR-CNEN Contract No. 012–61–12 BIAI. We wish to thank Dr. Raymond Appleyard for reading the manuscript.

SUMMARY

Theories were developed to predict the frequencies of various types of consanguineous marriages based on demographic data of migration patterns, age distributions, and similarity of mates in the general population. The effect of migration was formulated both with discrete and continuous models. In the former, the entire population is subdivided into discrete groups (villages etc.) and migration and marriage are treated using transition and matrimonial migration matrices. It was then shown that the method of matrix algebra leads to simple expressions of the results. However, information is not at the moment sufficient to construct numerically the migration and marriage matrices to treat the actual cases, but may become available in the future. On the other hand, using continuous models, fitting of either biexponential or trinormal distribution to migration distances allows us to predict the frequencies observed in an actual case (the Parma Valley area) when age effect on marriage is also taken into account.—The agreement between observed and expected results for Parma is only fair. In part,

at least, this seems to be the consequence of the inadequacy of the demographic information now available and that should be improved by future research.—As an indicator for the breeding structure of populations, the probabilities of consanguineous marriages should have an important bearing especially for human population genetics.

LITERATURE CITED

BARRAI, I., L. L. CAVALLI-SFORZA, and A. MORONI, 1962 Frequencies of pedigrees of consanguineous marriages and mating structure of the population. Ann. Hum. Genet. **25**: 347–377.

BARRAI, I., L. L. CAVALLI-SFORZA, and A. MORONI, 1964 Record linkage from parish books. pp. 51–60. *Mathematics and Computer Science in Biology and Medicine*, H.M.S.O., London.

BRAGLIA, G. L., 1962 Frequenze di matrimoni consanguinei. Tesi di Laurea in Fisica, Università di Parma.

CAVALLI-SFORZA, L. L., 1957 Some notes on the breeding patterns of human populations. Acta Genet. Statist. Med. **6**: 395–399. —— 1963 The distribution of migration distances: models, and applications to genetics. pp. 139–158. *Entretien de Monaco en Sciences Humanines; Les déplacements humains*. Edited by JEAN SUTTER.

CAVALLI-SFORZA, L. L., I. BARRAI, and A. W. F. EDWARDS, 1964 Analysis of human evolution and random genetic drift. Cold Spring Harbor Symp. Quant. Biol., **23**: 10–20.

FROTA-PESSOA, O., 1957 The estimation of the size of isolates based on census data. Am. J. Hum. Genet. **2**: 9–16.

HAJNAL, J., 1963 Random mating and the frequency of consanguineous marriages. Proc. Royal Soc. London B **159**: 125–177.

HALD, A., 1952 *Statistical Theory with Engineering Applications*. Wiley, London.

LUU-MAU-THANH, and J. SUTTER, 1963 Contribution à l'étude de la répartition des distances séparant les domiciles des époux dans un département francais. Influence de la consanguinité. Entretien de Monaco en sciences humaines; Les déplacements humains, 123–137.

MAINARDI, M., L. L. CAVALLI-SFORZA, and I. BARRAI, 1962 The distribution of the number of collateral relatives. Atti Ass. Genet. Ital. **7**: 123–130. —— 1966 Some demographic estimates of genetic interest. (in preparation).

MORTON, N. E., 1955 Non-randomness in consanguineous marriage. Ann. Hum. Genet., **20**: 116–124.

MORONI, A., Inbreeding explosion in the 19th Century in a Catholic country. (In preparation).

SUTTER, J., and TRAN NGOC TOAN, 1957 The problem of the structure of isolates and of their evolution among human populations. Cold Spring Harbor Symp. Quant. Biol., **22**: 379–383.

WRIGHT, S., 1943 Isolation by distance. Genetics **28**: 114–138.