

12-2-2008

## Chapter 2, Population Numbers

Alan C. Swedlund  
*Prescott College*

Follow this and additional works at: [https://scholarworks.umass.edu/anthro\\_res\\_rpt7](https://scholarworks.umass.edu/anthro_res_rpt7)



Part of the [Anthropology Commons](#)

---

Swedlund, Alan C., "Chapter 2, Population Numbers" (2008). *Research Report 07: The Genetic Structure of an Historical Population: a Study of Marriage and Fertility in Old Deerfield, Massachusetts*. 4.

Retrieved from [https://scholarworks.umass.edu/anthro\\_res\\_rpt7/4](https://scholarworks.umass.edu/anthro_res_rpt7/4)

This Article is brought to you for free and open access by the Anthropology Department Research Reports series at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Research Report 07: The Genetic Structure of an Historical Population: a Study of Marriage and Fertility in Old Deerfield, Massachusetts by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

## CHAPTER II

### POPULATION NUMBERS

#### Introduction

In developing theories of population genetics, the tendency in the past has been to assume that population size is infinite or, if finite, constant. In addition, problems with studying genetic structure have arisen not only because models are inadequate, but also because our knowledge of actual human populations has been deficient (Schull and MacCluer, 1968:282-83). It is quite clear that infinite population size is unrealistic for human population models, and a constant size is probably invalid in a number of specific, empirical situations. Although there are these problems in the concept and definition of population numbers, seldom do studies undertake clarification of the problems involved. In the following chapter the nature of this problem will be investigated in regard to the Deerfield records.

#### Effective Population Size

In attempting to characterize the genetic structure of human breeding populations, two variables are very commonly investigated: one is the effective population size (e.g. Wright, 1938; Kimura and Crow, 1963) and the other is the coefficient of inbreeding (e.g. Wright, 1931; Crow and Mange, 1965). These measures estimate departures from idealized conditions in the subject population. In a specific, localized, human population mating may not be random;

family size may vary greatly; and the population is not likely to be infinite. The effective population size is a parameter for defining these deviations from Hardy-Weinberg conditions.

The effective population size ( $N_e$ ) is "...the size of an idealized population that would have the same amount of inbreeding or of random gene frequency drift as the population under consideration" (Kimura and Crow, 1963:279), that is, under panmixia, the number producing the conditions observed in the subject population. As Falconer (1960:70) points out, probably the most common and important deviation from the system of an idealized population is the non-random distribution of family size. Formulae have been developed to estimate variations in family size. Wright (1938) presents a formula given constant population size:

$$N_e = \frac{4N - 2}{V + \bar{k}}$$

where  $N$  is the breeding population size,  $V$  is the variance in family size, and  $\bar{k}$  is the mean family size surviving to maturity; in populations of constant size this is equal to 2. Others (e.g. Kimura and Crow, 1963) have extended this to deal with separate sexes and varying population size.

Among human beings it is necessary to define what is meant by the breeding population, since parents and adults are not necessarily synonymous. Lasker (1954) and others have used the measure of parents with children at a given census time. This estimate can be high if it includes older, non-fertile parents, or low if it omits separated parents (Lasker, 1954:355). Others (e.g. Salzano et al, 1967) have defined the breeding population as composed of those individuals of reproductive age.

This latter definition has been adopted for the present study, and the breeding population is considered to be comprised of all those individuals between the ages of 16-45. This allows for the reality that individuals a "generation" apart may produce viable offspring and yet eliminates all parents who would normally be beyond reproductive age.

It is also necessary to emphasize that variance in family size (V) means variability in number of offspring who themselves reach maturity. This last consideration is very important, since sub-adult mortality could increase or reduce the variability observed at birth. The variance in family size for the Deerfield population was determined by taking all individuals who were parents in the year 1810, counting their total number of offspring, and then determining the mean and deviation in numbers of offspring for the total sample. 1810 was chosen because it appeared to be in a period of typical reproductive habits for Deerfield, and by taking parents at this time it was possible to include females who gave birth as early as 1789 and as late as 1837. This would compensate for possible fluctuations in social variables, disease, etc. Table 2.1 presents the basic information on the sample.

It may be noted that the family size for Deerfield at this time is very high, even for children surviving to the age of 16; the mean period of productivity for females is 15.9 years. While these values appear quite high, they are not inconsistent with values from other populations (Table 2.2).

The effective population size has been investigated in a few

TABLE 2.1  
REPRODUCTIVE HISTORY: PARENTS OF 1810

	All Children	Children Reaching 16 Yrs.
Number of families	41	40
Range of children	3-15	3-11
Mean Children	8.41	6.65
Std. Deviation	2.78	2.45
Variance	7.73	6.00
Mean Reproductive Period*	15.9 yrs.	15.9 yrs.
Deerfield Parents	65	63
Outside Parents	17	17

d=2.288      P < .01

\*Females

TABLE 2.2  
MEAN FERTILITY IN VARIOUS POPULATIONS

Population	Time	N	Source
Deerfield	c. 1810	8.41	Present Study
Plymouth Colony	c. 1700	8.56	Demos, 1965
U.S. Women	c. 1839	5.50	Crow & Morton, 1955
Hutterites	c. 1925	10.90	Henry, 1961*
Norway	c. 1875	8.10	"
Hindu Villages	c. 1945	6.20	"

\*In Spuhler, 1963.

human populations, and comparisons have been made between man and other animals. Crow and Morton (1955) calculated  $N_e$  for man, Drosophila, and the snail Lymnaea and found it to be between .70 and .95 of breeding size. Morton (1969:57) states that  $N_e$  for human females may be typically about two-thirds of breeding size. However, in Crow and Morton's study they considered mean family size at maturity to be two in all three species, since this is often found to be the case in natural populations (p. 211). It is my contention that this is not typical for many human "natural" populations. Even though population growth cannot go unchecked in any environment indefinitely, the fact is that the last 8,000 years of man's evolution have taken place under conditions of rapid increase, the rate of increase is most marked in the last few hundred years (see Huxley, 1956: Deevey, 1960). This trend has no doubt had microgeographic and microevolutionary significance as well as broader effects. As an example, the population of Deerfield grew to 5 times its original size in less than one hundred years, and doubled itself three times in its first 150 years (Figure 1.1). Migration certainly does not account for all this growth and large family size must be a contributing factor. This suggests then that constant population size is an unrealistic assumption for Deerfield and probably for the recent "natural history" of man.

It would thus appear that Crow and Morton (1955) may be too conservative in using the value of two for many human situations, and the results of an increase in mean family size and variance values are two-fold: (1) as mean family size increases, the size of the breeding population and effective population size also increase through time. The

reason for this is simply that large mean family size ultimately increases the absolute size of all fractions of the population by insuring that each generation will be larger than that preceding.

(2) As the mean family size and variance increase the relative proportion of effective size to breeding size decreases at a given point in time. That is, if a particular breeding population is the product of a family size and variance exceeding two, then the proportion of the effective population to breeding population will be less than if the population were not experiencing growth. This latter point is particularly important when investigating actual human populations. For example, the breeding population (individuals between 16-45) of Deerfield in 1810 consisted of 649 individuals. If the population is considered constant in size over time then the effective size is 563 using Kimura and Crow's (1963) formula:

$$N_e = \frac{4N-4}{V^*+2}$$

where  $V^*$  is the variance for one sex (females=2.6) and  $N$  is breeding population size. However, if the actual mean (6.65) is used, and a constant rate of growth, but not size, is assumed, then the effective size becomes 107 using Kimura and Crow's (1963) general formula:

$$N_e = \frac{N_t - 2\bar{k} - 2}{\bar{k} - 1 + V^*/\bar{k}}$$

where  $N_t - 2$  is the grandparental generation size which, given a constant rate of growth is equal to  $N_t - 1 / \bar{k}$ . In a species with separate sexes, such as man, a pair of alleles in an individual

cannot come from the same parent, or from two individuals of the same sex. A pair of alleles may come, however, from the same grandparent. The difference between the two formulae above is that, if population size is constant, the parental and grandparental generation are the same; but if population size is changing, then the  $N$  of the grandparents should be used, and the mean ( $\bar{K}$ ) will be greater than 2.

Table 2.3 presents the effective population estimates for Deerfield using the assumed and observed values. The estimates include the effective population size given a mean family size of two, and, in addition, the values given for actual mean family size. The rather dramatic differences between the values are apparent, and relevant in regard to the fact that several past studies have used mean family size of two when other values were observed (e.g. Lasker, 1954; Kuchemann et al, 1967; Salzano et al, 1967). Although Deerfield is an extreme example in the sense that family size is so large, it is indicative of the direction and magnitude in which  $N_e$  may vary.

TABLE 2.3

EFFECTIVE POPULATION SIZE BASED ON ASSUMED  
AND OBSERVED FAMILY SIZE: DEERFIELD, 1810

Total Population Size=1570; Breeding Population=649

	V	X	$N_e$	%N*	%T
Assumed	2.6	2.00	563	86.7	35.8
Observed	2.6	6.65	107	16.5	6.8

\*N=breeding size; T=total size



The effective population size, as presented on the previous page, has a further weakness in the sense that it is based on a strict generational construct. This presents itself in calculation as the myth that 6.65 children, in the case of Deerfield, occur as a single event which all fertile females share in common. As mentioned above, at any given time (e.g. 1810) females bearing young may have also borne children 20 years before or after, and the breeding population is in a constant state of change. The complexity which overlapping generations creates is not easily dealt with in man (Schull and MacCluer, 1968). Kimura and Crow (1963) have defined  $N_e$  for overlapping generations with constant population size:

$$N_e = \frac{N^2}{N_0 r} = \frac{N}{b r}$$

where  $N$  is the total population number,  $N_0$  is the number born per unit time,  $b = N_0/N$  is the crude birthrate, and  $r$  is the average age of reproduction. Again, the problem arises in populations undergoing growth. In populations which have not reached stability and where age will vary with time, effective population size cannot be viewed as a stable relative proportion of the population. The effective population size will change relatively and absolutely.

A second, very important variable which, although difficult to measure, will affect the effective population size is migration (Morton, 1969:57). Most measures of effective population size are based on the concept of an idealized situation in which no migration is occurring. Lasker (1954) states that in "primitive" or "folk" cultures the breeding population is more or less synonymous with the community

(p. 353). It would appear that "less synonymous" may be most appropriate, since he then goes on to state that over 20 percent of the parents he analyzed in Paracho in 1952 were from outside the community. In calculating  $N_e$  it is necessary either to assume that the effects of immigration and emigration are equal, in numbers as well as genotypes, or to make some effort to estimate possible differences. Since  $N_e$  is intended to define the sampling variance in gene frequencies between parents and offspring, it is probably most correct to accept the migration existing in the parent group, and to make adjustments for migration by altering the denominator of the equation. Thus, if migration is a factor, then in addition to adjusting mean family size to reflect those who survive to maturity, it is also necessary to account for those who will be gained or lost through migration. For example, if emigration is reducing the number of individuals reaching maturity in the local population, then the rate of this emigration should be added to the rate of mortality between birth and maturity in determining mean family size.

In addition to changing population size and migration, other factors will have an impact on the effective population size (Salzano et al, 1967:488): (1) concentration of relatives in the founding group; (2) restriction of mate selection within the population; and, (3) differential inheritance of fertility. The imprecisions which attach to effective population size thus become manifold. This has led Morton and Yasuda (1962:188) to state that: "Because of its mathematical simplicity, the concept of a subpopulation with an assignable size  $N$  has fascinated population geneticists to such an extent as to retard the development of a more realistic theory."

If a specific community happens to be the focal point of a genetic study, then in spite of the difficulties in quantifying a genetically significant measure of size, some indication of the changes in size of the local breeding population can be useful. Changes in the size will reflect: (1) the growth or decline of the genetically significant reproductive portion of the population; (2) changes in the age structure of the subject population; and, (3) the effects of migration and mortality on the population when viewed through time. The difficulty arises in determining which of these three may be causing any fluctuations observed. Figure 2.1 presents the relative and absolute growth of the breeding population of Deerfield through time. The size of the breeding population appears to be on the increase relatively as well as absolutely. The increase is probably attributable to both high local fertility and immigration, but as indicated above, this high local fertility would have the effect of decreasing the relative effective size of the population.

#### Coefficient of Inbreeding

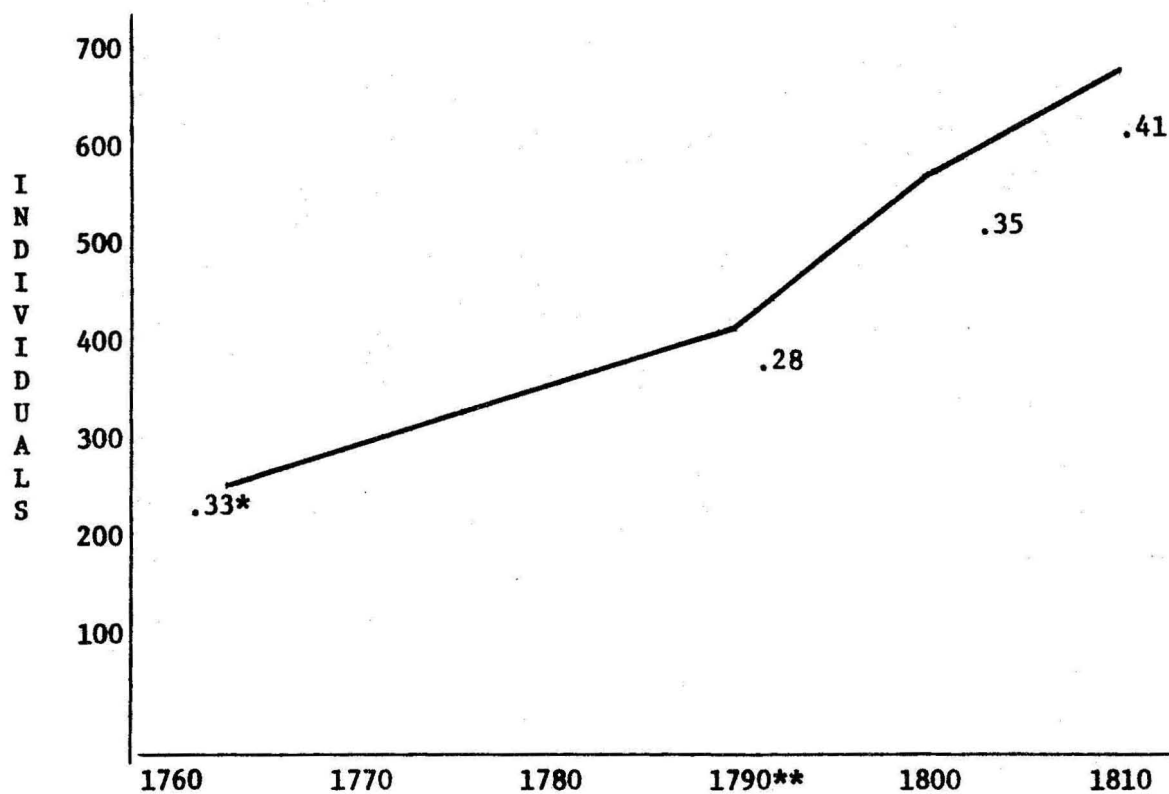
As discussed above, the effective population size is an estimate which ultimately is an expression of inbreeding and gene drift. Inbreeding ( $F$ ) may be defined as the mating together of individuals related by ancestry. The coefficient of inbreeding is the probability that two genes at any locus in an individual are identical by descent (Falconer, 1960:60-61).

Inbreeding has two components, the random component, which is a sampling product of small population size, and indicative of the

FIGURE 2.1

RELATIVE AND ABSOLUTE GROWTH OF THE BREEDING POPULATION OF DEERFIELD: 1765-1810

Individuals 16-45 years



\*Proportion of breeding size to total population.

\*\*Values had to be extrapolated from records not recording exact age breakdown.

opportunity for drift; and a nonrandom component, which in human beings is the tendency for related individuals to marry. Numerous formulae have been developed to estimate inbreeding under various conditions. The most common of these is the model for analysis of individual pedigrees:

$$F = \sum \left[ \left(\frac{1}{2}\right)^{n_1+n_2+1} (1+F_A) \right]$$

Wright (1922), where  $n_1$  is the number of generations from one parent back to the common ancestor and  $n_2$  from the other parent, and  $F_A$  is the inbreeding coefficient of the common ancestor.

One estimation of inbreeding which has been developed for human populations and which can be used for subpopulations where migration occurs is based on the frequency of isonomic marriages (Crow and Mange, 1965). This estimate of inbreeding has recently been applied to several populations and, while caution is warranted regarding the fact that surnames are not genes, isonomy has shown reasonable agreement with other estimates based on European data (Yasuda and Morton, 1967; Morton, 1969). The principle behind the calculation of inbreeding by isonomy is an assumption that all isonomy is a reflection of common ancestry. "Let  $F$  be the total inbreeding coefficient,  $F_r$  be the inbreeding from random mating within the population, and  $F_n$  be that from nonrandom marriages. These are related by

$$F = F_n + (1 - F_n) F_r,$$

where

$$F_n = (P - \sum p_i q_i) / 4 (1 - \sum p_i q_i)$$

and

$$F_r = \sum p_i q_i / 4$$

approximately" (Crow and Mange, 1965:201). Where  $p_i$  is the proportion of males with a certain name,  $q_i$  is the corresponding proportion in females, and  $P$  is the proportion of isonomic marriage pairs.

The Deerfield marriage records were analyzed for inbreeding by the above model. A total of four samples were drawn: (1) a sample including all marriages in Deerfield,  $N=1470$ , (2) a sample including all endogamous marriages in Deerfield,  $N=734$ , and, time based samples for marriages occurring between (3) 1790-1809,  $N=633$ , and, (4) 1820-1839,  $N=677$ . The results are summarized in Table 2.4.

TABLE 2.4  
INBREEDING ESTIMATED BY ISONOMY, DEERFIELD RECORDS

Sample		$I^*$	$F_r$	$F_n$	$F$
Total	1470	.0177	.00207	.00233	.00433
Endogamous	734	.0191	.00273	.00202	.00474
1790-1809	633	.0063	.00045	.00110	.00155
1820-1839	677	.0118	.00055	.00242	.00295

$I^*$ =Isonomy frequency

The overall conclusion to be reached from these data (Table 2.4) is that marriage in Deerfield has not been significantly different from random; however, the values also indicate changes in expected directions. For example, the coefficients for endogamous marriages are higher than those for all marriages except for the non-random

component. This may indicate a preference for marriage with relatives outside the local community, which has been found to be the case in other populations studied (Morton, 1964; Freire-Maia and Freire-Maia, 1962). The time-based samples also show that inbreeding tends to increase through time. This has been found for other sub-populations (e.g. Hutterities, Yasuda and Morton, 1967) and is an indirect confirmation of the nature of population growth discussed above; that is, large family size would tend to increase the likelihood of relatives marrying each other and thus to decrease the relative effective population size. Thus for Deerfield and other growing populations the localized factors tend to mitigate against a large proportional effective population size. Under the above conditions, elevation of the effective population size will be attained only by migration.

One interesting aspect of the present study is that it is possible to trace the reproductive performance of isonymous pairs and determine whether or not close inbreeding has any notable effects on fertility. Of the total of 26 isonymous pairs, 18 are found to have some biographical information available, the remainder either emigrated at marriage (4 cases), or no information was available (4 cases). Fifteen of the 18 are known cousin pairs, and 12 include what could be considered complete fertility information (of the remaining, two spouses had died within a year of marriage and one had moved away after four years of marriage). The 12 known pairs range from first cousin to second cousin-once-removed matings.

The mean completed family size for cousin marriages ( $N=N-1$ ) is markedly below that of the females who were parents in 1810 (Table 2.5).

TABLE 2.5  
MEAN COMPLETED FAMILY SIZE OF KNOWN CONSANGUINEOUS  
MARRIAGES AND THE PARENTS OF 1810

Sample	$N_p$	$N_o$	$\bar{X}$	Sigma
Consanguineous	12	36	3.27	2.78
Parents of 1810	41	345	8.41	2.78

$N_p$  = parent pairs       $N_o$  = offspring

Using a t-test of significance the differences are highly significant between the two means ( $t=5.44$ ,  $d.f.=50$ ,  $P<.001$ ). These values do not include postnatal mortality which would presumably be higher in consanguineous matings.

Although it is possible to estimate the amount of inbreeding in a human population such as Deerfield, we find that an estimate of population size is very difficult, and perhaps meaningless. Even though the mating pattern in Deerfield is essentially random, there is no close similarity between this community and the isolate or neighborhood model in human genetics. On the other hand, marriage tends to be most frequent among community residents and those in the nearest neighboring communities--so that mating is not entirely random



over distances greater than the community. The fact that migration is an important factor in the genetic structure of Dearfield, and presumably most communities, and that patterns of mating and fertility will be affected by migration, requires some means of expressing this significant mechanism.

In the Introduction reference was made to the fact that when a human community is being studied, as opposed to other communities of animals, the possible effects of culture must be considered. In the foregoing discussion it is important to take note of the fact that mating with neighboring communities may be based upon, or may tend to establish, important cultural ties. These ties may, in turn, reinforce interbreeding between neighboring communities. This process will have the effect of increasing the likelihood of inbreeding among individuals in these communities

The attempt in this chapter to define the concept and problem of population numbers leads to the observation that although numbers are very important to an understanding of genetic structure, population size is, at the very best, difficult to quantify. Since population numbers are so closely related to the nature of movements of people, perhaps statements of probability concerning migration are the best form of estimate.