Estimation of fitness reduction due to a chronic disease in man

BY M. NEI

Center for Demographic and Population Genetics, University of Texas at Houston,
Houston, Texas 77025

AND S. YOKOYAMA

Department of Psychiatry, Washington University School of Medicine and the Jewish Hospital of
St Louis, 216 South Kingshighway, St Louis, Missouri 63110

SUMMARY

An improved mathematical method is presented for estimating the amount of fitness reduction due to a chronic disease by using demographic data. It is shown that Cavalli-Sforza and Bodmer’s equivalent formula gives an underestimate. Application of the new formula indicates that the selective difference between blood groups O and A, resulting from their association with duodenal ulcer, is $6.4 \times 10^{-5}$, i.e. ten times higher than Cavalli-Sforza and Bodmer’s estimate.

INTRODUCTION

Cavalli-Sforza & Bodmer (1971, p. 336) introduced a method for estimating the amount of fitness reduction due to a chronic or complex genetic disease such as duodenal ulcer, using demographic data of the disease. Theoretically, this method is applicable to any disease whether the genetic basis of the disease is known or not. Unfortunately, their formula does not seem to be appropriate (Nei, 1973), since they ignored the incidence of the disease in the population in computing the fitness reduction. In order to evaluate this quantity it is essential to divide the population into two groups, i.e. the individuals (trait group) who will eventually develop the disease (or trait) and those (non-trait group) who will not. Of course, it is not always easy to know the frequency of the trait group at birth. When data on ages at onset are available, however, it is possible to estimate this frequency and thus the amount of fitness reduction. In this note we shall present one such method. Before going into the detail, however, we would like to emphasize that the amount of fitness reduction estimated by the following method is not the same as the usual selection coefficient in population genetics unless the disease is controlled by a single Mendelian gene with complete penetrance. Nevertheless, it seems to be useful to know the quantity, as will be discussed later.

MATHEMATICAL FORMULATION

Let $y_a$ be the proportion of individuals in the trait group who contract the disease at age $a$. Clearly, $1 - \Sigma_{a=0}^{x} y_a$ is the proportion of individuals in the trait group who are still normal at the end of age $x$. We assume that if the trait is developed, the individual may die due either to the trait or to other causes. Let $q_x^t$ and $q_x^z$ be the probabilities of death due to the trait and other causes at age $x$, respectively. Assuming the independence of deaths due to the trait and
other causes, the probability of survival of an individual with the trait at age \( x \) is given by \((1 - q^T_x)\) \((1 - q^L_x)\). The equivalent survival probability for individuals without the trait is of course \((1 - q^C_x)\).

It is clear that the probability of survival up to the end of age \( x \) for the non-trait group is

\[
I_x^0 \equiv \prod_{t=0}^{x} (1 - q^C_t).
\]  

On the other hand, the probability of survival up to the end of age \( x \) for an individual who developed the trait at age \( a \) is given by

\[
I_{a,x} \equiv \prod_{t=0}^{a-1} (1 - q^T_t) \prod_{t=a}^{x} (1 - q^L_t)(1 - q^C_t)
= I_x^0 \prod_{t=a}^{x} (1 - q^L_t),
\]  

where \( a \leq x \). Therefore, the probability of survival up to the end of age \( x \) for the trait group is

\[
I_x^1 \equiv \left(1 - \sum_{a=0}^{x} y_a\right) I_x^0 + \sum_{a=0}^{x} y_a I_{a,x}.
\]  

Hence, the average survival probability for the total population is

\[
I_x = pI_x^0 + (1 - p)I_x^1,
\]  

where \( p \) is the proportion of the trait group at birth.

Following Cavalli-Sforza and Bodmer, we measure fitness in terms of the rate of reproduction \((R)\). Let \( b_x^T \) and \( b_x^0 \) be the age-specific birth rates for the trait group and non-trait group, respectively, and \( b_x \) be the birth rate for the total population. Therefore,

\[
b_x = pb_x^T + (1 - p) b_x^0.
\]  

The reproductive rate for the non-trait group is given by

\[
R^0 = \sum_{x=0}^{m} I_x^0 b_x^0,
\]  

whereas that for the trait group is

\[
R^1 = \sum_{x=0}^{m} (1 - Y_x) I_x^0 b_x^0 + \sum_{x=0}^{m} b_x^1 \sum_{a=0}^{x} y_a I_{a,x},
\]  

where \( m \) is the maximum age at which children can be produced and \( Y_x = \sum_{a=0}^{x} y_a \). Therefore, the average reproductive rate for the total population is

\[
R = pR^1 + (1 - p)R^0.
\]  

If we assume that the generation time is the same for the trait and non-trait groups, the amount of fitness reduction \((\delta)\) due to the trait may be obtained by

\[
\delta = (R^0 - R^1)/R^0
= \frac{1}{R^0} \sum_{x=0}^{m} \left[ Y_x b_x^0 - b_x^1 \sum_{a=0}^{x} y_a \prod_{t=a}^{x} (1 - q^L_t) \right].
\]
Estimation of fitness reduction due to chronic disease

It is therefore clear that in order to estimate \( s \) information on \( p_x^0, b_x^0, y_a, b_x^1 \), and \( q_i^1 \) is necessary.

The above formula is different from Bodmer's (1968) equivalent formula. The difference occurred because Bodmer considered the probability of developing the trait at age \( x \) rather than the proportion of individuals in the trait group who develop the trait at age \( x \) and his methods of formulation and approximation were somewhat different from ours.

When the individuals in the trait group become infertile as soon as they develop the disease, \( b_x^1 = 0. \) In this case \( s \) may be estimated if \( p_x^0, \ Y_x, \) and \( b_x^0 \) are known. On the other hand, when \( b_x^1 = b_x^0 \) but the individual with the trait has an increased mortality, \( s \) is given by

\[
s = \frac{1}{R^0} \sum_{x=0}^{m} p_x b_x \left[ Y_x - \sum_{a=0}^{x} y_a \prod_{t=a}^{x} (1-q_i^1) \right]. \tag{10}\]

In practice, separate demographic data for the trait and non-trait groups are not generally available. When \( p \) is small, however, \( p_x^0, b_x^0, \) and \( R^0 \) may be approximated by the values (\( l_x, b_x, \) and \( R \)) for the total population. Therefore,

\[
s = \frac{1}{R} \sum_{x=0}^{m} l_x b_x \left[ Y_x - \sum_{a=0}^{x} y_a \prod_{t=a}^{x} (1-q_i^1) \right] \tag{10'}
\]

approximately. It is also noted that when \( b_x^1 \) is not known, the upper bound and lower bound of \( s \) may be estimated by assuming \( b_x^1 = 0 \) and \( b_x^1 = b_x^0, \) respectively.

In some diseases the trait is expressed in the early stage of life, well before reproductive age. In this case (9) reduces to

\[
s = \frac{1}{R^0} \sum_{x=0}^{m} l_x \left[ b_x^0 - b_x^1 \prod_{t=0}^{x} (1-q_i^1) \right]. \tag{11}\]

When \( q_i^1 \) is small compared with 1 and \( b_x^1 = b_x^0, \) (11) may be approximated by

\[
s = \frac{1}{R^0} \sum_{x=0}^{m} p_x b_x \sum_{t=0}^{x} q_i^1. \tag{12}\]

DISCUSSION

Cavalli-Sforza & Bodmer (1971) derived a formula for estimating fitness reduction, which, in our terminology, becomes \( s' = (R^0 - R)/R. \) They obtained this formula since they did not really separate the trait and non-trait groups. If we note \( R = pR^1 + (1-p)R^0 = R^0 + p(R^1 - R^0), \) \( s \) in (9) is given by

\[
s = (R^0 - R)/(pR^0). \tag{13}\]

Therefore, when \( R^0 \) is close to \( R, \) \( s = s'/p \) approximately. Thus, Cavalli-Sforza and Bodmer's formula is expected to give a serious underestimate of \( s \) when \( p \) is small. In their study of the fitness reduction due to duodenal ulcer Cavalli-Sforza and Bodmer used the formula

\[
s' = \frac{1}{R} \sum_{x=0}^{m} l_x b_x \sum_{t=0}^{x} q_i^t, \tag{14}\]

where \( q_i^t = k_t q_t, \) in which \( q_t \) is the age-specific mortality for the total population and \( k_t \) is the proportion of all deaths at age \( t \) that are caused by duodenal ulcer. (Their formula on page 337 involves some apparent misprints.) This has the same form as that of (12). (Cavalli-Sforza and
Bodmer implicitly assumed that duodenal ulcer is expressed in the early stage of life.) However, \( q'_t \) is different from \( q_t \), which is simply the proportion of deaths due to the trait at age \( t \) in the trait group (not in the total population). Namely, \( q'_t = pq_t \). Therefore, when \( p \) is small, there is a large difference between \( q'_t \) and \( q_t \), and consequently between the estimates \( s \) and \( s' \).

Cavalli-Sforza & Bodmer (1971) used Italian data on vital statistics for females in 1953 for estimating the fitness reduction due to duodenal ulcer. According to them (table 6.6 in their book), the values of \( q'_t \) for the age groups 15–19, 20–24, 25–29, 30–34, 35–39, 40–44 and 45–49 are 5.8, 8.15, 10.2, 15.5, 20.0, 39.7 and 77.7 respectively, all being multiplied by \( 10^{-6} \). Using these and the values of \( l_x \) and \( b_x \) for the same age groups, they obtained \( s' = 3.85 \times 10^{-5} \) through (14). However, since only a small proportion of people contract duodenal ulcer, this is certainly an underestimate.

There seems to be no data on \( y_x \), \( l_x \), \( b_x \), and \( q'_x \) for ulcer patients. However, if we assume \( l'_x = l_x \), \( b'_x = b_x \), \( R'_x = R \), and accept Cavalli-Sforza and Bodmer's assumption that ulcer is expressed before reproductive age, then it is possible to estimate \( s \) by (12). In the case of duodenal ulcer these assumptions do not appear to lead to any serious error, since the rate of death due to this disease is very small. According to Silen (1977), about 10% of the population suffers from duodenal ulcer, so that \( p = 0.1 \). Therefore, \( q'_t \) is \( 10q_t \), and this leads to the estimate of \( s = 3.85 \times 10^{-4} \). Namely, the amount of fitness reduction due to duodenal ulcer seems to be ten times larger than Cavalli-Sforza and Bodmer's estimate.

In this note we defined \( p \) as the proportion of individuals at birth who will eventually develop the trait. In general, this is not measurable at the time of birth. However, if the proportion of individuals who develop the disease at each age group is known, \( p \) may be estimated by the sum of the proportion over all age groups.

As mentioned earlier, the \( s \) value in this note is not necessarily equal to selection coefficient. Many genetic diseases in man do not have complete penetrance and are often controlled by many loci. In these cases \( s \) is obviously an overestimate of selection coefficient. For example, when a disease is controlled by a dominant gene with penetrance \( x \), the actual selection coefficient against the dominant heterozygote would be \( 2x \) rather than \( s \). Furthermore, even if the disease considered is totally environmental, the above method would generally give a positive value of \( s \). Therefore, caution should be exercised in the interpretation of the result obtained. Nevertheless, the above method is useful for estimating the amount of selection resulting from a chronic disease, whether the disease is genetic or not. Furthermore, when a chronic disease is associated with a particular genotype at a polymorphic locus, it is possible to estimate the contribution of the disease to the selection against the genotype (Cavalli-Sforza & Bodmer, 1971).

For example, blood group O is susceptible to duodenal ulcer 1.4 times more often than blood group A. From this information, Cavalli-Sforza and Bodmer estimated that the selective difference between blood groups O and A that is due to duodenal ulcer is \( 6.4 \times 10^{-5} \) for females. This estimate is based on their value of \( s' = 3.85 \times 10^{-6} \). If we use our improved estimate of fitness reduction due to duodenal ulcer, it becomes \( 6.4 \times 10^{-5} \). At any rate, this type of information would be important for understanding the mechanism of maintenance of genetic polymorphism.

This study was supported by National Institutes of Health research grants GM 19543 and GM 20293.
Estimation of fitness reduction due to chronic disease

REFERENCES


