Am J Hum Genet 25:598–603, 1973

Total Number of Individuals Affected by a Single Sex-linked Deleterious Mutation in a Finite Population

WEN-HSIUNG LI[1]

A deleterious mutation may cause pathological and clinical consequences extending over several generations. A recent catalog of inherited abnormalities by McKusick [1] lists hundreds of distinct clinical syndromes, each of which can be plausibly attributed to the effect of single-locus mutations. From the public health viewpoint it is important to know both the total number of individuals affected by a single deleterious mutation before it is eliminated from the population and its extinction time. In the case of autosomal genes, these two quantities have recently been studied by Nei and Li [2–4] for various types of mutations and population sizes. However, their results are not directly applicable to sex-linked deleterious mutations. The purpose of this paper is to extend the results to sex-linked genes.

TOTAL NUMBER OF MUTANT INDIVIDUALS

Consider a population in which a mutant gene a and its wild-type allele A are segregating at a sex-linked locus. Let the relative fitnesses of genotypes in males and females be as specified in table 1. The selection coefficients against mutant males (s) and homozygous mutant females (s') are positive constants but the selection coefficient (h') for the heterozygous females can be negative. (However, we will be concerned only with h' such that $s + 2h'$ is positive and not extremely small.) We assume that no homozygous mutant female occurs in the whole process. This is reasonable since we are considering only deleterious mutations in fairly large populations. We also assume equal sex ratio and equal gene frequency in both sexes so that the genotype frequencies are as specified in table 1 for a given mutant gene frequency x in a random mating population. Strictly speaking, the gene frequency is not the same for males and females if there is selection, and there is really no justification given of the fact that we can make a diffusion approximation to this two-variable non-Markovian process using only one overall gene frequency (x). However, it will be seen later that these assumptions do not affect the final result seriously. In the following we will consider a randomly mating population of actual size N and

Received March 1, 1973; revised May 4, 1973.

Paper no. 1623 from the Laboratory of Genetics, University of Wisconsin. Supported by grant no. GM 15422 from the National Institutes of Health.

[1] Department of Medical Genetics, University of Wisconsin, Madison, Wisconsin 53706. Present address: Center for Demographic and Population Genetics, University of Texas, Houston, Texas 77025.

TABLE 1

RELATIVE GENOTYPE FITNESS AND FREQUENCY

GENOTYPE	MALE		FEMALE		
	A	*a*	*AA*	*Aa*	*aa*
Fitness	1	$1-s$	1	$1-h'$	$1-s'$
Frequency	$1-x$	x	$1-2x$	$2x$	0

effective size N_e. Note that the effective population size for a sex-linked gene is different from that for an autosomal gene (see [5, 6] for explanation).

Let x be the frequency of a and $\phi(p, x; t)$ be the probability density of x at generation t, given that $x = p$ at $t = 0$. Mathematically, the total numbers of heterozygous females and mutant males can be treated in the same way. Let $f(x)$ be either the number of heterozygous females or that of mutant males in a population. Explicitly, $f(x) = Nx$ for the number of female heterozygotes and $f(x) = Nx/2$ for that of mutant males, under the above assumptions. The total number of female mutant heterozygotes or mutant males before the mutant gene becomes extinct is given by

$$n_f(p) = \int_0^\infty \int_0^1 f(x)\, \phi(p,x;t)\, dxdt. \tag{1}$$

It can be shown (see [4]) that $n_f(p)$ satisfies the equation

$$\frac{V_{\delta p}}{2} n_f''(p) + M_{\delta p}\, n_f'(p) = -f(p), \tag{2}$$

with boundary conditions $n_f(0) = 0$ and $n_f'(1)$ finite. The $M_{\delta p}$ and $V_{\delta p}$ in equation (2) are the mean and variance of gene frequency change per generation, respectively. Approximately, $M_{\delta p} = -(s + 2h')p/3$ and $V_{\delta p} = p(1 - p)/2N_e$.

1. *Total number of female heterozygotes.* In this case $f(p) = Np$, as mentioned earlier. Therefore, equation (2) is satisfied by $(d/dp)(1 - p)^K n_1'(p) = -4N_eN(1 - p)^{K-1}$, where $K = 4N_e(s + 2h')/3$. After one integration, it becomes $n_1'(p) = 3N/(s + 2h') + [C_1 - 3N/(s + 2h')]\ (1 - p)^{-K}$. The condition that $n_1'(1)$ is finite implies $C_1 = 3N/(s + 2h')$. Therefore we obtain

$$n_1(p) = 3Np/(s + 2h'), \tag{3}$$

since $n_1(0) = 0$. If initially there is only one mutant gene, then $p = 2/3N$ and

$$\bar{n}_1 \equiv n_1(2/3N) = 2/(s + 2h'). \tag{4}$$

If the mutant gene is completely dominant ($h' = s$), formula (4) reduces to

$$\bar{n}_1 = 2/3s. \tag{5}$$

2. *Total number of mutant males.* By putting $f(p) = Np/2$ in equation (2), we can readily obtain

$$n_2(p) = 3Np/2(s + 2h'). \tag{6}$$

If initially there is only one mutant, then

$$\bar{n}_2 \equiv n_2(2/3N) = 1/(s + 2h'). \tag{7}$$

If the mutant gene is completely recessive ($h' = 0$), then formula (7) becomes

$$\bar{n}_2 = 1/s. \tag{8}$$

This is exactly as expected since in this case selection against the mutant gene occurs only in males. If the mutant is completely dominant ($h' = s$), then

$$\bar{n}_2 = 1/3s, \tag{9}$$

and the total number of individuals affected is

$$\bar{n}_1 + \bar{n}_2 = 1/s. \tag{10}$$

This is also exactly as expected, since the solution is essentially the same as for an autosomal locus [4].

It is interesting to note that $\bar{n}_1$ and $\bar{n}_2$ are independent of population size for any type of mutation and that the overdominant effect of a mutant gene is very small for any population size. These results are quite contrary to the results for autosomal mutations and will be discussed in more detail later.

AVERAGE EXTINCTION TIME

It has been shown by Kimura and Ohta [7] that the expected extinction time $T(p)$ satisfies

$$\frac{V_{\delta p}}{2} T''(p) + M_{\delta p} T'(p) = -U_0(p), \tag{11}$$

with boundary conditions $T(0) = 0$ and $T'(1)$ finite, where $U_0(p)$ is the probability of eventual extinction. Since we are concerned only with sex-linked deleterious mutations in relatively large populations, there is almost no probability for the mutant to be fixed in the population, and we can assume that $U_0(p) = 1$ and the average extinction time $\bar{t}_0(p) = T(p)/U_0(p) = T(p)$, if p is not large. Due to the linearity of equation (11) in $T(p)$, it can be shown that the solution of equation (11) is equivalent to $T(p) = T_1(p) + T_2(p)$, where

$$T_1''(p) - \frac{K}{1-p} T_1'(p) = -\frac{4N_e U_0(p)}{1-p} \tag{12}$$

and

$$T_2''(p) - \frac{K}{1-p} T_2'(p) = -\frac{4N_e U_0(p)}{p}. \tag{13}$$

It can be easily shown that $T_1(p) \leqslant 3p/(s + 2h')$, which is negligible since we consider only small initial mutant frequencies. From equation (13), we have

$$T_2'(p) = 4N_e(1 - p)^{-K} \int_p^1 x^{-1}(1 - x)^K U_0(x)dx + C_1(1 - p)^{-K}.$$

That $T_2'(1)$ is finite implies $C_1 = 0$. Therefore we obtain

$$T_2(p) = 4N_e \int_0^p (1 - y)^{-K}dy \int_y^1 x^{-1}(1 - x)^K U_0(x)dx,$$

since $T_2(0) = 0$. The integrand $x^{-1}(1 - x)^K U_0(x)$ can be roughly replaced by $x^{-1}(1 - x)^K$, since $U_0(x)$ is almost 1 when x is not large and $x^{-1}(1 - x)^K$ is very small when x is large. Thus the average extinction time is given roughly by

$$\bar{t}_0(p) = T_2(p) = 4N_e \int_0^p (1 - y)^{-K}dy \int_y^1 x^{-1}(1 - x)^K dx. \qquad (14)$$

The double integration in the above formula is complicated but it can be shown that the influence of population size on $\bar{t}_0$ is very small compared with the case of autosomal mutations. For simplicity, let us consider one extreme case, say, $h' = -s/2$. Then $K = 4N_e(s + 2h')/3 = 0$ and $\bar{t}_0(2/3N) = 8N_e[1 + \ln(3N/2)]/3N$, which increases very slowly with increasing N. It can be shown that $\bar{t}_0$ is generally smaller than the above expression for any given N.

DISCUSSION

The results by diffusion approximations have been seen to be exactly as expected if the mutant gene is completely recessive or dominant [see formulas (8) and (10)]. However, in order to see whether diffusion methods are also adequate for the other cases, let us study a discrete time model. It is expected that discrete time models give better results than continuous time models when the selection coefficient is large. Consider a lethal mutation with overdominant effect, that is, $s = 1$ and $h' < 0$. If the mutation is first expressed in a male, obviously only one person will be affected. On the other hand, if it occurs first in a female, in the following successive generations the numbers of mutant males will be $(1 - h')/2$, $(1 - h')^2/4, \ldots$. Therefore the total number of males affected in this case will be $(1 - h')/(1 + h')$. Assuming mutation occurs with probability $1/3$ in males and $2/3$ in females, the average total number of mutant males is therefore $1/3 + 2(1 - h')/3(1 + h') = (1 - h'/3)/(1 + h')$. The difference between the result by the diffusion method and that by a discrete model is therefore $d = -2h'(1 - h')/[3(1 + h')(1 + 2h')]$. This is generally very small. To illustrate, take a rather extreme case, say $h' = -0.20$ (20% overdominance); then $d = 1/9$, which is rather small. Therefore the results by diffusion methods are seen to be simple and reasonably accurate.

In this paper we have assumed that the population size remains constant. In many human populations, however, the size is increasing. In an increasing population the total number of individuals affected by a single mutation will be larger and the extinction time will be longer than those given in this paper. However, if the

absolute fitness of mutant males in the increasing population is less than one, the formulas for the total numbers of individuals affected given in this paper are still applicable if we take absolute values for $1 - s$ and $1 - h'$. At any rate, our computations may serve as minimum estimates if the population size continues to increase. Of course, human populations cannot increase indefinitely and in the future they would reach a steady state or fluctuate around some "optimal" number.

In some genetic diseases such as hemophilia, the classical concept of recurrent mutation appears to hold. In this case it is important to know the expected frequencies of female mutant heterozygotes and mutant males at steady state in finite populations. The expected relative frequency of heterozygotes among females is given by

$$\overline{f_1} = (2N_f + N_m)u\,\tilde{n}_1/N_f = 2(2N_f + N_m)u/N_f(s + 2h'), \tag{15}$$

and that of mutant individuals among males by

$$\overline{f_2} = (2N_f + N_m)u\,\tilde{n}_2/N_m = (2N_f + N_m)u/N_m(s + 2h'), \tag{16}$$

where N_f and N_m are the numbers of females and males in the population, respectively, and u is the mutation rate per locus per generation, assuming equal mutation rate in both sexes. These two formulas can be derived by following the method of Li and Nei [4].

It is interesting to compare the mutational damage caused by a sex-linked mutation and by an autosomal mutation. If the mutant gene is completely dominant, then these two quantities are the same and equal to $1/s$. This is because the death of an individual eliminates one gene, regardless of whether it is sex-linked. If the mutant gene is completely recessive then $\tilde{n}_2$ equals $1/s$ for a sex-linked mutation but $1/2s$ for an autosomal mutation. This is because in the former case death occurs only in mutant males and one death eliminates only one gene, while in the latter case death occurs in mutant homozygotes and each death eliminates two genes. If the mutant gene shows overdominant effect, then $\tilde{n}_2$ equals $1/(s + 2h')$ for a sex-linked mutation but

$$[1 - \sqrt{\pi}\,A\,e^{A^2}\,\mathrm{erfc}(A)]/[2(s - 2h')]$$

for an autosomal mutation as obtained by Li and Nei [4], where

$$A = h'\sqrt{2N_e(s - 2h')}/s$$

in the present notation. Note that the former is independent of population size while the latter is highly dependent on population size. When population size is small, the difference between these two quantities is relatively small, but in a large population the difference may be extremely large. For example, if $h' = -0.02$ and $s = 1$, then $\tilde{n}_2$ is only 1.04 for a sex-linked gene regardless of population size, while it is 3.4 for an autosomal mutation if $N_e = N = 1{,}000$ but 10,000 if $N_e = N = 10{,}000$. This is because an autosomal deleterious mutation with overdominant

effect may persist in the population for a very long time if the population size is large, while a sex-linked deleterious mutation can survive only for a short period of time even if the population is very large. To illustrate this, let us consider an extreme case. For a sex-linked mutation with $h' = -0.25$ and $s = 0.5$, $\overline{t_0}$ is only 16.6 generations if $N = 1{,}000$ and $N_e = 3N/4$. But for an autosomal mutation with $h' = -0.25$ and $s = 1$, $\overline{t_0}$ is 5.1×10^{35} generations if $N_e = N = 1{,}000$, as was computed by Li and Nei [4]. Therefore, a sex-linked overdominant deleterious mutation is much less harmful to a society than an autosomal overdominant deleterious mutation.

SUMMARY

The expected total numbers of heterozygous females ($\bar{n}_1$) and mutant males ($\bar{n}_2$) affected by a sex-linked deleterious mutation and the average extinction time ($\overline{t_0}$) are studied using diffusion methods. In sharp contrast to autosomal mutations [4], $\bar{n}_1$ and $\bar{n}_2$ are independent of population size, and the effect of population size on $\overline{t_0}$ is small regardless of the degree of dominance. The effects of overdominance on these quantities are also very small compared with the case of autosomal mutations. When a mutation is dominant, the total number of individuals affected for a sex-linked locus is the same as that for an autosomal locus. If a mutation is completely recessive, the former is twice as large as the latter. For an overdominant mutation, the former is much less than the latter, particular in large populations.

ACKNOWLEDGMENT

I would like to thank Professors James F. Crow, Masatoshi Nei, and Sewall Wright for many discussions and valuable comments.

REFERENCES

1. McKusick VA: *Mendelian Inheritance in Man. Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes,* 3d ed. Baltimore, Johns Hopkins Press, 1971
2. Nei M: Total number of individuals affected by a single deleterious mutation in large populations. *Theor Pop Biol* 2:426–430, 1971
3. Nei, M: Extinction time of deleterious mutant genes in large populations. *Theor Pop Biol* 2:419–425, 1971
4. Li WH, Nei M: Total number of individuals affected by a single deleterious mutation in a finite population. *Am J Hum Genet* 24:667–679, 1972
5. Wright S: *Evolution and the Genetics of Populations, vol 2: The Theory of Gene Frequencies.* Chicago, Univ. Chicago Press, 1969
6. Crow JF, Kimura M: *An Introduction to Population Genetics Theory.* New York, Harper & Row, 1970
7. Kimura M, Ohta T: The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61:763–771, 1969