Environmental Health Perspectives
Vol. 52, pp. 41-44, 1983

Are Human Studies Possible? Some Thoughts on the Mutation Component and Population Monitoring

by Carter Denniston*

The concept of the mutation component of a genetic trait is discussed and its relation to heritability investigated. It is found that for qualitative traits held in the population by opposing directional selection and mutation, the mutation component equals the broad sense heritability. The mutational component of a quantitative trait with an intermediate optimum is found generally to be between half and one times the (narrow sense) heritability of the character. Although more work on this matter is required, the finding of a relationship between mutation component and heritability should allow us to make better predictions regarding the likely impact of an increased mutation rate on the human population. Some statistical problems relevant to the detection of an increased mutation rate are also discussed.

Committees attempting to assess the effect of an increased mutation rate on the human population always face the difficulty posed by what BEIR III (*1*) called "irregularly inherited" disorders. They state:

> "The population survey of British Columbia reported that at least 9% of live born humans will be seriously handicapped at some time during their life times by genetic disorders of complex etiology, manifested as congenital malformations, anomalies expressed later, or constitutional and degenerative diseases. This, the largest category of genetic disorder . . . we refer to as 'irregularly inherited' disorders."

They go on:

> "An estimate of the number of induced irregularly inherited disorders present at equilibrium must take into account the proportion of the incidence of these disorders that would vary directly with the mutation rate, a quantity that BEIR I called the 'mutational component.' More precisely, if the equilibrium incidence, I, of a disorder is a linear function of the mutation rate, m, i.e., $I = a + bm$, then we define the mutational component to be $MC = bm/(a + bm)$, in which case the relative increase of the disorder incidence after an increase in the mutation rate from m to say, $m(1 + k)$ is $(I' - I)/I = (MC)k$. Each disorder may have its own mutational component, and a class of disorders, such as irregularly inherited disorders, its average mutational component."

I would like to discuss the idea of "mutation component" today; it may turn out to be a more useful concept than we realized when the report was written.

Crow and Denniston (*2*) define the mutation component M_i of a disorder i, such that its impact changes from I_i to $I_i(1 + M_i k_i)$ when the mutation rate changes from μ_i to $\mu_i(1 + k_i)$. Then the total impact, $I = \Sigma I_i$, of a group of disorders changes from I to $I(1 + M\tilde{k})$ when the total mutation rate changes from U to $U(1 + \tilde{k})$, where $U = \Sigma\mu_i$ is the total mutation rate, $\tilde{k} = \Sigma\mu_i k_i/\Sigma\mu_i$ is the average (mutation weighted) increment; $M = \Sigma I_i k_i M_i/I\tilde{k}$ is the total mutation component.

From these definitions we find that the mutation component for disorder i is

$$M_i = (\mu_i/I_i)(\Delta I_i/\Delta\mu_i)$$

For small increments in the mutation rate, it may be convenient to express this definition in terms of differentials so that $M_i = d\ln I/d\ln \mu$, showing its relation to the concept of elasticity common in economic theory.

The "impact" of a disorder is meant to refer to its detrimental effect on human well-being, measured simply as incidence, load or perhaps some more relevant index of human suffering. I shall assume, here, that such an index is proportional to incidence or decrease in fitness; the proportion-

*Department of Genetics and Medical Genetics, University of Wisconsin, Madison, WI 53706.

Table 1.

Genotype	AA	AB	AB	BB
Frequency	p^2	$2pq\pi$	$2pq(1 - \pi)$	q^2
Fitness	$1 - s$	$1 - s$	1	1

ality contant may, of course, vary from one disorder to another.

For simple monogenic disorders, the calculation of mutation component is straightforward. Consider an autosomal locus segregating two alleles, A and B (Table 1), where π is a measure of incomplete penetrance or incomplete dominance. We then have that the gene frequency at equilibrium between mutation and selection is given by

$$\hat{p}^2 s(1 - 2\pi) + \hat{p}s\pi(1 + \mu) - \mu = 0$$

where μ is the rate of mutation from B to A. If $\pi = 0$, A is recessive and $\hat{p}^2 = \mu/s$; if $\pi = 1$, A is dominant and $2\hat{p} \simeq 2\mu/s$; more generally, $2\hat{p} \simeq \mu/s\pi$, as long as selection is acting primarily on heterozygotes. In any event the mutation load ranges between μ and 2μ (*3*). We see that both the incidence and load of such disorders, and therefore their impact on human well-being, are directly proportional to the mutation rate. The mutation component is thus equal to one. If all cases are not of genetic origin, then the incidence is of the form $a + b\mu$, where a represents the phenocopies. The mutation component is $b\mu/(a + b\mu)$, which corresponds to the broad sense heritability of the disorder if one considers it as a 0, 1 trait.

It appears that for a large class of disorders held in the population by opposing directional selection and mutation, the mutation component is equal to the broad sense heritability (*2*). This conclusion can be upset by genotype × environment interactions and, possibly, by arbitrary gene interactions and linkage; but enough special cases have been looked at to suggest fairly wide applicability.

It is often the case that an equilibrium frequency is a function of more than one mutation rate, e.g., at an X-linked locus with mutation rates differing in eggs and sperm, when the trait under consideration involves more than one gene, or if a mutagen affects forward and backward mutation rates differently. We may extend our concept of mutation component as follows.

We define the "partial mutation component" of a trait relative to mutation rate j as $M_j = (\mu_j/I)\partial I/\partial\mu_j$. With this definition when

$$\mu_j \rightarrow \mu_j(1 + k_j)$$

we have

$$\begin{aligned}\Delta I/I &= \Sigma\,(\partial I/\partial\mu_j)\mu_j k_j/I \\ &= \Sigma\, M_i k_i\end{aligned}$$

For example, the equilibrium frequency for a recessive X-linked trait is approximately

$$I = (2m_1 + m_2)/s$$

where m_1 and m_2 are the mutation rates in eggs and sperm, respectively, and s is the selection coefficient. If the mutation rates are changed in the proportions k_1 and k_2, then

$$M_1 = 2m_1/(2m_1 + m_2)$$

and

$$M_2' = m_2/(2m_1 + m_2)$$

so that

$$\begin{aligned}\Delta I/I &= M_1 k_1 + M_2 k_2 \\ &= (2m_1 k_1 + m_2 k_2)/(2m_1 + m_2)\end{aligned}$$

Disorders held in the population by opposing selection forces, e.g., sickle cell anemia, are not expected to respond to an increase in the mutation rate. For example, the mutation component of a trait like sickle cell anemia is $M \simeq 2m/sq$, where m is the mutation rate to the sickle cell gene, $1 - s$ is the fitness of the normal homozygote and q is the equilibrium gene frequency of the S allele. With intermediate values of q, as long as s is substantially larger than m, we see that M is quite small.

Many traits, e.g., blood pressure, weight, are continuous with an optimum and large deviations from that optimum being highly disadvantageous. What is the mutation component of these traits? The problem is a difficult one, but we may make some progress by looking at a specific model. Kimura (*4*) has constructed such a model. He assumes a quantitative character under the control of many loci at each of which is an infinite number of alleles acting additively with respect to the character. There is an optimal phenotype with respect to fitness and fitness decreases in proportion to the squared deviation from this optimum. Under these assumptions, the genetic load is given by

$$L = u^{1/2}A + uB^2 + KV_E$$

where u is the common mutation rate, $A = \Sigma(2K\bar{x}_i^2)^{1/2}$, $B = \Sigma\bar{x}_i/(\bar{x}_i^2)^{1/2}$; here, $\bar{x}_i$ is the average effect of a mutation at the ith locus and $\bar{x}_i^2$ is the average squared effect; K is a measure of the intensity of selection and V_E is the environmental variance. The mutation component of the load is then

$$M_L = \frac{[(1 + k)^{1/2} + 1]^{-1}A + u^{1/2}B^2}{(A/h^2 + u^{1/2}B^2)}$$

where k is the increment in mutation rate and h^2 is the narrow heritability. If $B = 0$, i.e., mutation produces deviations in both directions about equally frequently, we find that

$$M_L = [(1 + k)^{1/2} + 1]^{-1} h^2$$

so that for small k the mutation component is equal to half the narrow sense heritability. This same result is derived by Crow and Denniston (*2*) under somewhat more general assumptions. Their conclusion is "for a measured trait determined by additive genes and independent environmental effects, where the fitness (and therefore the impact) is proportional to the squared deviation from the optimum and where selection is weak, the mutation component for small changes in the mutation rate is between ½ and 1 times the heritability. If the mean is close to the optimum, the factor is close to ½; if the mean is far from the optimum, the factor becomes larger and approaches 1 as a limit when selection is entirely directional.

The model for a quantitative trait is more restrictive than that for a qualitative one, and it has been necessary to resort to more approximations. The conjecture is that, since the conclusions for directional selection are robust with respect to the way the genetic factors interact, this might also be true for quantitative traits with an intermediate optimum, but this remains to be seen as further studies along the lines begun by Kimura and Lande (*5–7*) are done.

So for qualitative disorders, if the narrow heritability is high (and therefore so is the broad heritability), the mutation component is large; if the broad heritability is low (and therefore so is the narrow heritability), the mutation component is small. If the broad heritability is large and the narrow heritability is small, the mutation component is indeterminate, but the response to an increase in the mutation rate is likely to be very slow so that the impact on the human population will be spread out over many generations. This can be seen by considering the meaning of a low narrow heritability. If the narrow heritability is low, it means either the environmental variance is large or much of the genetic variance is due to genetic interactions, dominance and epistasis. The result is that a parent's phenotype is less predictive of his child's, the trait shows an irregular pattern of inheritance, and the genes responsible are often hidden from the scrutiny of selection. The approach to the new equilibrium following an increase in the mutation rate will be exceedingly slow and most of the damage will occur far in the future. In fact, if, for example, improvements in medical treatment keep pace, no discernable damage may occur at all.

The mutation component of a quantitative polygenic trait with its mean near the optimum is approximately half the narrow heritability of the quantitative character.

So far I have discussed the problem of predicting the likely effect on the population of an increased mutation rate. Equally important is the detection of such an increase, if and when it occurs. This is an interesting statistical problem primarily because of the "when." We don't know ahead of time when the increase will occur.

The situation is this: We sample from the population at regular intervals (e.g., monthly, quarterly, yearly) and observe the spontaneous frequency of some trait with a high mutation component. We wish to detect any shift (usually an upward shift) in this spontaneous frequency as quickly as possible, so that its cause can be investigated and possibly corrected. On the other hand, we do not want to cry wolf too often. Of course, if the shift has already occurred and we are told when it occurred, the problem is a standard one; we merely compare two samples, before and after the shift. However, our problem is a sequential one. Two approaches have been discussed in the literature and my intention here is simply to point them out and suggest that some work should be done to determine which is the more appropriate to our needs.

The statistical situation may be formulated as follows: Given observations on the independent random variables $X_1, X_2, \ldots, X_n$ (taken at consecutive times) which are distributed according to the distribution function $F(X:\Theta_i)$, $i = 1, \ldots, n$, we wish to test the simple hypothesis H_0: $\Theta_1 = \ldots = \Theta_n = \Theta_0$ (Θ_0 known) against the composite alternative H_1: $\Theta_1 = \ldots = \Theta_m = \Theta_0$; $\Theta_{m+1} = \ldots = \Theta_n = \Theta_0 + \Delta$, where both the change point m and the magnitude of the change Δ are unknown. And we want to do our testing sequentially in such a way as to minimize the average time to rejection of H_0 after a shift has occurred and maximize the average time to rejection of H_0 in the absence of any shift, subject to the constraints of sample size, money and so on.

Morton (*8*), in an interesting article, has suggested using Wald's sequential probability ratio method for human population monitoring. He demonstrates the method using data on Down's syndrome from Australia and Sweden. His application is essentially a two-sided sequential test, assumes an underlying Poisson distribution, and tests the simple hypothesis $\Theta = \Theta_0$ against the simple alternative $\Theta = K\Theta_0$. He discusses the

relations between sample size, K and the expected number of samples required for a decision for Type I and Type II error rates of 10%. Morton states his preference for the Wald scheme to the cumulative sum method (discussed below), citing greater flexibility and ease of interpretation. Nevertheless, it seems unlikely that the Wald scheme could be optimal since it is not specifically directed at the alternative hypothesis of interest, i.e., a shift in the parameter during the sampling process.

The other approach, advocated by Weatherall and Haskey (*9*), comes out of the continuous inspection schemes used for quality control in industry and pioneered by Page (*10*) and Barnard (*11*). In its simplest form, the cumulative sum method consists of recording the cumulative sums

$$S_r = \sum_{i=1}^{r} (X_i - k)$$

where k is often the expected value of X under H_0. S_r is set to zero if it becomes negative. H_0 is rejected when the cumulative sum reaches some predetermined value, h. Again, one wants to detect a shift quickly but not often claim a shift when none has occurred; to this end one wants the expected value of r at rejection to be small under H_1 and large under H_0. The relevant relations between h, k, sample size and these expected run lengths are discussed, for the Poisson and normal distributions, by Ewan and Kemp (*12*). A deep discussion of the general change point problem has been given by Kander and Zacks (*13*).

It is important that we improve our ability to predict the likely impact of an increased mutation rate and that we measure that impact in ways useful to decision makers. If our predictions are reliable and alarming, society would be advised to expend resources on controlling mutation rates; if our predictions are unreliable or suggest little immediate impact, resources are probably better spent on other pressing social problems. It is hoped that a development of the concept of mutation component may help us in these predictions. The other problem, of detecting an increase in the mutation rate when it occurs, is of equal importance and was discussed briefly from a statistical point of view.

REFERENCES

1. National Research Council, Advisory Committee on the Biological Effects of Ionizing Radiation. The Effects on Populations of Exposure to Low Levels of Ionizing Radiation. National Academy of Sciences, Washington DC, 1980, pp. 85, 86.
2. Crow, J. F., and Denniston, C. The mutation component of genetic damage. Science 212: 888–893 (1981).
3. Haldane, J. B. S. The effect of variation on fitness. Am. Naturalist 71: 337–49 (1937).
4. Kimura, M. A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. 54: 731–736 (1965).
5. Lande, R. The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. 26: 221–235 (1976).
6. Lande, R. The influence of the mating system on the maintenance of genetic variability in polygenic characters. Genetics 86: 485–493 (1977).
7. Lande, R. The genetic covariance between characters maintained by pleiotropic mutations. Genetics 94: 203–205 (1980).
8. Morton, N. E. Surveillance of Down's Syndrome as a paradigm of population monitoring. Human Hered. 26: 360–371 (1976).
9. Weatherall, J. A. C., and Haskey, J. C. Surveillance of malformations. Brit. Med. Bull. 32: 39–44 (1976).
10. Page. E. S. Continuous inspection schemes. Biometrika 41: 100–116 (1954).
11. Barnard, G. A. Control charts and stochastic processes. J. Roy. Statist. Soc. B 21: 239–257 (1959).
12. Ewan, W. D., and Kemp, W. D. Sampling inspection of continuous processes with no autocorrelation between successive results. Biometrika 47: 363–380 (1960).
12. Kander, Z., and Zacks, S. Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. Ann. Math. Statist. 37: 1196–1210 (1966).