Title
ON THE INTERPRETATION OF THE DOSE FREQUENCY RADIO-GENETICS

Author
Wijsman, R.A.

Publication Date
1951-08-29

Reprinted from GENETICS, Vol. 36, No. 5, September, 1951; pages 478–487.

ON THE INTERPRETATION OF THE DOSE-FREQUENCY CURVE IN RADIOGENETICS*

R. A. WIJSMAN[1]

Department of Physics, University of California, Berkeley, California

Received November 17, 1950

IN an article under the same title by OPATOWSKI (1950) and in another by OPATOWSKI and CHRISTIANSEN (1950) doubt has been raised whether a linear relationship between mutation frequency and irradiated dose, at low dosage, can be considered as proof that a mutation is induced by one event. The conclusion reached in these papers is that under the opposite assumption, *i.e.*, that a mutation needs a large number of primary events, it is quite possible to get a frequency-dose curve which is essentially linear in the range in which measurements have been made (SPENCER and STERN 1948). Since this conclusion seems at first sight rather strange, it was considered worth while to make a more thorough analysis of the problem.

The purpose of this paper is to investigate in several cases whether or not the theoretical mutation frequency curve can be distinguished from a single event-type curve, which is essentially a straight line in the lower range of the curve. The question how well the experimental data justify a straight line fit, is of an entirely different nature and will not be considered here. Since the experimental data cover a mutation frequency range from about 0.001 to 0.1, the theoretical curves will be investigated in this same range.

Our calculations will be based on the following model: A certain mutation is a certain chemical change in a particular gene, effected by one or more ionizations or excitations in that gene. An ionization or excitation will henceforth be called an "event." We suppose that the events are independent of each other. Let n be the number of events which have to occur in order to produce the mutation, and let N be the total number of events which possibly can take place in the gene, so that we have $n \leqq N$. Since the gene may be a big molecule, N may be very large. However, we do not want to include in N any possible events which cannot contribute to the mutation under consideration. Therefore, even if the gene is a big molecule the number N is not necessarily large (it may be even 1). We shall distinguish between two cases: 1) $N \gg 1$, and 2) N not $\gg 1$. In case 1) we let N approach infinity in order to simplify the calculations. In both cases we distinguish further between the two possibilities: a) all events have the same probability of occurrence; b) the events may have different probabilities. In each case the mutation frequency is the probability that out of N possible events at least n events actually occur. The model described above does not cover all possible cases, but it covers a range sufficiently wide to be of interest.

* Part of the cost of the mathematical formulas has been paid by the U. S. ATOMIC ENERGY COMMISSION.

[1] PUBLIC HEALTH SERVICE RESEARCH FELLOW OF THE NATIONAL CANCER INSTITUTE.

DISCLAIMER

The existence of spontaneous mutations is no complication in the present treatment, since, following OPATOWSKI, we can consider these spontaneous mutations as being caused by a fictitious radiation of dose S. If we denote the dosage of the actual radiation by D, the total effective dose is D + S.

After application of this effective dose the probability that the i^{th} event has not taken place is $q_i = \exp[-c_i(D+S)]$, in which c_i is a constant. It follows that the probability p_i that the i^{th} event has taken place is

$$p_i = 1 - q_i = 1 - \exp[-c_i(D+S)].$$

For convenience we introduce the following dimensionless quantities

$$c = \sum_{i=1}^{N} c_i, \quad f_i = \frac{c_i}{c},$$

and:

$$x = c(D+S) \tag{1}$$

so that we have:

$$\sum_{i=1}^{N} f_i = 1 \tag{2}$$

$$q_i = \exp[-f_i x] \tag{3}$$

$$p_i = 1 - \exp[-f_i x] \tag{4}$$

We now proceed to discuss the four different cases 1a, 1b, 2a and 2b separately.

Case 1a: ($N \to \infty$, f_i is the same for all i).

From (2) follows $f_i = 1/N$ and substitution into (4) gives $p_i = p = 1 - \exp\left[-\frac{x}{N}\right]$. The average number of events is

$$\lim_{N\to\infty} Np = \lim_{N\to\infty} N\left\{1 - \exp\left[-\frac{x}{N}\right]\right\} = x,$$

which can be seen most easily after expanding the exponential in a power series. Now in our case, since $N \to \infty$, the probability of exactly m events is given by Poisson's formula:

$$P_m = \frac{1}{m!} e^{-x} x^m \tag{5}$$

If at least n events are necessary for the mutation, then the probability M_n of the mutation is $M_n = 1 - \sum_{m=0}^{n-1} P_m$, or, after substitution of (5):

$$M_n = 1 - \sum_{m=0}^{n-1} \frac{1}{m!} e^{-x} x^m \tag{6}$$

which also can be written in the form:

$$M_n = \frac{1}{(n-1)!} \int_0^x y^{n-1} e^{-y} dy \tag{7}$$

That eqs. (6) and (7) are equivalent can be seen by integrating (7) by parts n times.

M_n can be plotted as a function of x for various values of n. For small n formula (6) is convenient, for large n formula (7) can be used to investigate the shape of the mutation frequency curve in the limit of $n \to \infty$. In taking this limit we may replace $n-1$ in (7) by n:

$$M_\infty = \lim_{n\to\infty} \frac{1}{n!} \int_0^x y^n e^{-y} dy \tag{8}$$

If n is large compared with 1 we can approximate the integrand in (8) by a Gaussian function, *i.e.*, a function of the form $A \exp[-B(y-C)^2]$, with A, B and C positive constants, and choose these constants such as to let the two curves have the same maximum and the same curvature at the maximum. The maximum of $y^n e^{-y}$ lies at $y = n$, and the value at the maximum is, of course, $n^n e^{-n}$, so $C = n$ and $A = n^n e^{-n}$. The second derivative of $y^n e^{-y}$ at $y = n$ is $-n^{n-1} e^{-n}$; the second derivative of $n^n e^{-n} \exp[-B(y-n)^2]$ at $y = n$ is $-2B n^n e^{-n}$. Equating the second derivatives of the two functions gives $B = \frac{1}{2}n$. We have, finally:

$$y^n e^{-y} \cong n^n e^{-n} \exp\left[-\frac{1}{2n}(y-n)^2\right], \quad (n \gg 1) \tag{9}$$

For $n \gg 1$ this function has a sharp peak at $y = n$, and if we take in (8) the lower limit of the integral $-\infty$ instead of 0, the error we make approaches zero as n approaches infinity. Thus we have:

$$M_\infty = \lim_{n\to\infty} \frac{1}{n!}\, n^n e^{-n} \int_{-\infty}^x \exp\left[-\frac{1}{2n}(y-n)^2\right] dy.$$

Using Stirling's approximation:

$$n! \cong n^n e^{-n} \sqrt{2\pi n}, \quad (n \gg 1) \tag{10}$$

and denoting by $H(x)$ the function:

$$H(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi$$

we can put M_∞ in the form:

$$M_\infty = \lim_{n\to\infty} \tfrac{1}{2}\left\{1-H\left(\frac{n-x}{\sqrt{2n}}\right)\right\} \tag{11}$$

It has been stated before that we take the mutation frequency of spontaneous mutations as $M = 0.001$. Let this value be assumed in $x = x_0$, then x_0 corresponds to $D = 0$. If we substitute $M = 0.001$ and $x = x_0$ into (11) we can solve for the argument of the function H. Using the Tables of the W. P. A. PROJECT FOR THE CITY OF NEW YORK (1941) we obtain: $\lim_{n\to\infty} \frac{n-x_0}{\sqrt{2n}} = 2.185$. We now put $\lim_{n\to\infty} \frac{x-x_0}{\sqrt{2n}} = z$, so that $\lim_{n\to\infty} \frac{n-x}{\sqrt{2n}} = 2.185 - z$. It is important to note that z is a variable proportional to D. In terms of the new variable (11) becomes:

$$M_\infty = \tfrac{1}{2}\{1-H(2.185-z)\}, \quad (z \geqq 0) \tag{12}$$

By plotting M_∞ against z one obtains the shape of the mutation frequency curve, with $n \to \infty$, for values of $M \geqq 0.001$.

The results so far obtained have been plotted in figure 1 with solid lines. Curve *a* is M_1, curve *b* is M_2, both curves calculated from (6): curve *c* is M_∞, calculated from (12). The curves for $n = 3$, 4, etc., not drawn in the figure, lie between curves *b* and *c*. For the sake of comparison the curves have been shifted and compressed in horizontal direction in such a way as to make them intersect at $M = 0.001$ and $M = 0.1$.

It is easily seen from figure 1 that the only curve which is essentially linear in the range considered is curve *a, i.e.,* the curve based on a single event model. The curve for $n = 2$ deviates conspicuously from a straight line, and the deviation becomes greater with increasing n.

Case 1b: ($N \to \infty$, f_i arbitrary).

As in case 1a we denote by P_m the probability of exactly m events, and by M_n the probability of at least n events, so that we still have $M_n = 1 - \sum_{m=0}^{n-1} P_m$. However, because of the arbitrariness of the f_i, P_m is not necessarily given by Poisson's formula (5), and, therefore, the expression (6) for M_n may not hold in this case. We shall investigate under which conditions the equations (5) and (6) will still be valid.

The probability P_0 of no events at all is the product of the probabilities q_i, given by (3):

$$P_0 = \lim_{N\to\infty} \prod_{i=1}^{N} q_i = \lim_{N\to\infty} \prod_{i=1}^{N} \exp[-f_i x] = e^{-x}$$

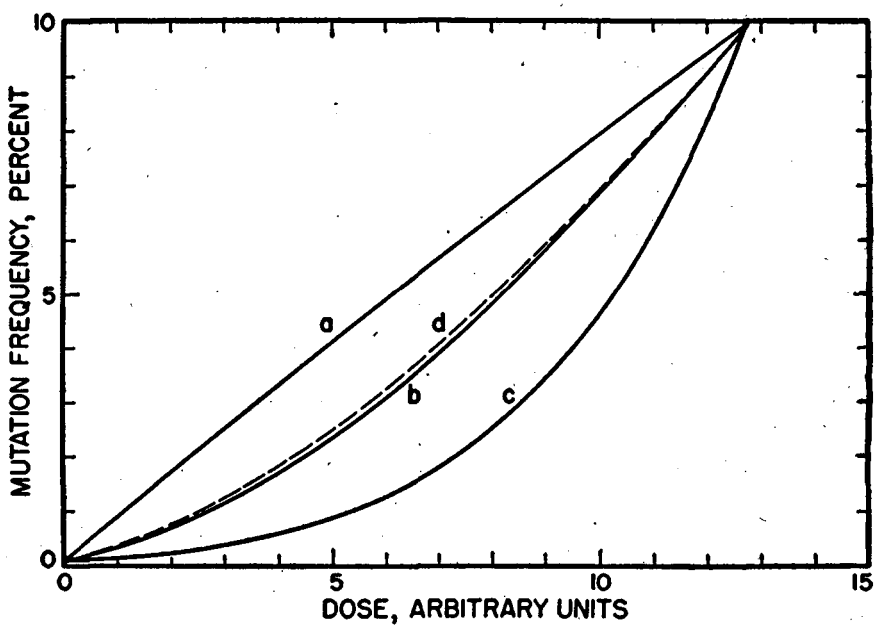

FIGURE 1.—Mutation frequency M as a function of the radiation dose D, in the frequency range $0.001 \leqq M \leqq 0.1$, for several values of n and N. Curve *a*: n = 1, N arbitrary; curve *b*: n = 2, N = ∞; curve *c*: n = ∞, N = ∞; curve *d*: n = 2, N = 2. All other values of n and N give curves lying between the curves *d* and *c*.

The single event curve *a* is compared with the multiple event curves *b*, *c* and *d*. The single event curve is approximately a straight line, whereas the multiple event curves are obviously curved.

in view of (2). This is the same expression as given by Poisson's formula (5), with m = 0, so that the mutation frequency for n = 1, $M_1 = 1 - e^{-x}$, is the same as in case 1a. This result is independent of special assumptions concerning the f_i.

The probability of exactly one event is

$$P_1 = \lim_{N\to\infty} \sum_{i=1}^{N} p_i \prod_{j\neq i} q_j = \lim_{N\to\infty} \sum_{i=1}^{N} (1-\exp[-f_i x])\exp[(-1+f_i)x],$$

which can be brought into the form:

$$P_1 = e^{-x} \lim_{N\to\infty} \sum_{i=1}^{N} (e^{f_i x} - 1) \tag{13}$$

After expanding the exponentials under the summation sign, (13) can be written

$$P_1 = e^{-x} \lim_{N\to\infty} \sum_{i=1}^{N} (f_i x + \tfrac{1}{2} f_i^2 x^2 + \ldots) \tag{14}$$

It is impossible to examine the shape of P_1 unless we make some assumption about the f_i. The physically most plausible assumption is that, although the f_i may be quite different from each other, none of them will be very much larger than any of the others. In terms of our physical model this means that none of the events which can take place in the gene occurs with a much greater probability than any of the other events. Under this assumption, and in view of (2), it is clear that $\lim f_i = 0$ for each i, in such a way that

$$\lim_{N\to\infty} \sum_{i=1}^{N} f_i^m = 0 \text{ for } m \geq 2.$$

In (14) we keep only the terms of first order: $P_1 = e^{-x} \lim_{N\to\infty} \sum_{i=1}^{N} f_i x = xe^{-x}$, by virtue of (2). This is Poisson's formula (5) with $m = 1$. Thus we get for M_2 in case 1b the same expression as in case 1a, *i.e.*, (6) with $n = 2$. By the same type of argument it can be shown that, in general, P_m is given by (5), and, therefore, M_n is given by (6) just as though all the f_i were equal. Therefore, under the assumption made concerning the f_i, we reach the same conclusion as in case 1a.

If we do not assume that none of the f_i is extraordinarily large compared to any of the others, we must allow a possibility that, for instance, one of the f_i is large compared to the sum of all the other f_i. It can be shown that in this case an n-event model produces an $(n-1)$-event type curve, *i.e.*, a curve given by (6) after replacing n by $n-1$. Physically the reason for this is clear: if the radiation dose is sufficient to produce with appreciable probability one of the events with small f_i, the one event which is comparatively so much more probable will then have been produced simultaneously with almost certainty, so that the existence of this event has no influence on the shape of the mutation frequency curve. Similarly, if there are k events, each of which is much more probable than the other $N-k$ events together, the number characterizing the type of mutation frequency curve is reduced by k. In particular, if $n = k+1$, *i.e.*, the minimum number of events necessary to cause a mutation is one more than the number of exceptionally probable events, the mutation frequency curve will have the shape of a one event curve.

Since there does not seem to be direct evidence against the above mentioned possibilities, we cannot quite exclude them. However, these exceptional possibilities seem physically rather unreasonable and unlikely, and we shall not consider them any further.

Case 2a: $\left(N \text{ not} \gg 1,\ f_i = \frac{1}{N} = \text{same for all i}\right)$.

The difference between this case and the preceding cases again lies in the expression for the probability P_m of exactly m events. As in case 1a, the probabilities of occurrence and of not occurrence of the i^{th} event are

$p = 1 - \exp\left[-\frac{x}{N}\right]$ and $q = \exp\left[-\frac{x}{N}\right]$ respectively, independent of i. From this we derive:

$$P_m = \binom{N}{m} p^m q^{N-m} = \binom{N}{m} \left(1 - \exp\left[-\frac{x}{N}\right]\right)^m \left(\exp\left[-\frac{x}{N}\right]\right)^{N-m},$$

or

$$P_m = \binom{N}{m} e^{-x} (e^{\frac{x}{N}} - 1)^m \tag{15}$$

For $m = 0$ we still have $P_m = e^{-x}$, just as in the case $N \to \infty$ so that M_1 is still given by (6) with $n = 1$, representing a one event type curve. For P_1 we find from (15): $P_1 = e^{-x} N(e^{x/N} - 1)$, and after expanding the exponential we get:

$$M_2 = 1 - P_0 - P_1 = 1 - e^{-x} - xe^{-x}\left(1 + \frac{x}{2N} + \frac{x^2}{6N^2} + \ldots\right) \tag{16}$$

The bigger N is, the more M_2 approaches $1 - e^{-x} - xe^{-x}$, which is the expression for M_2 in the case 1a or b, and which was already shown to be non-linear in the pertinent range. In the worst case $N = 2$, in which we can write M_2 as:

$$M_2 = (1 - e^{-\frac{x}{2}})^2, \quad (N = 2) \tag{17}$$

This curve has been plotted in figure 1 as curve *d* (dotted line). Again the horizontal scale has been adjusted such as to facilitate comparison with the other curves. It is seen that curve *d* is very close to curve *b*, and can be distinguished very well from the one event curve *a*. Since the curves for $n = 2$ and $N > 2$ all lie between curves *b* and *d*, our conclusion holds for $n = 2$ and any N.

If one investigates the curves for $n = 3$ and various $N \geq 3$, it turns out that none of these curves is straighter than $(1 - e^{-x/3})^3$. Since the latter is curved more than $(1 - e^{-x/2})^2$, is follows that all curves for $n = 3$ and any N can be well distinguished from the one event curve M_1. The same conclusion holds for higher values of n.

Case 2b: (N not $\gg$ 1, f_i arbitrary)

Again we have M_1 given by (6) with $n = 1$. M_2 is equal to

$$M_2 = 1 - e^{-x} - xe^{-x}\left(1 + \tfrac{1}{2}x \sum_{i=1}^{N} f_i^2 + \ldots\right).$$

Under the same assumptions as made in case 1b, *i.e.*, that none of the f_i is very large compared with the other f_i, each f_i is of the order of magnitude $\frac{1}{N}$, and therefore $\sum_{i=1}^{N} f_i^2$ is of the order $N \cdot \frac{1}{N^2} = \frac{1}{N}$. We write $\sum_{i=1}^{N} f_i^2 = \frac{a}{N}$, in which *a* is a constant close to 1. Substitution into M_2 gives:

$$M_2 = 1 - e^{-x} - xe^{-x}\left(1 + a\frac{x}{2N} + \ldots\right) \tag{18}$$

Since the constant a is close to 1, and appears in front of a term which itself is a correction term, M_2 given by (18) is very close to M_2 given by (16), which was already shown to be appreciably curved.

It can be shown that also for higher values of n the expressions for M_n are well approximated by the corresponding expressions in case 2a, all of which yielded curves which could be well distinguished from the one event curve.

CONCLUSION AND COMPARISON WITH OPATOWSKI'S RESULTS

We conclude from the foregoing that, if our physical model is correct, a linear relationship between mutation frequency and radiation dose is a strong indication that a mutation is caused by a single event.

The question arises why the conclusions of this paper are so very different from those reached in the papers by OPATOWSKI and OPATOWSKI and CHRISTIANSEN. In order to answer this question we shall examine the equation derived in OPATOWSKI's paper, which is equation (3) in that, paper, and which will be denoted here by (O3):

$$P(D) = C \int_{-s}^{D} \exp\left[-h^2(x-M)^2\right] dx \qquad (O3)$$

In this equation D and S have the same meaning as in the present paper; P is the mutation frequency (which was called M throughout this paper); h and M are constants, the values of which can be chosen arbitrarily and independently; and C is determined by the normalization $P(\infty) = 1$.

The assumptions underlying OPATOWSKI's derivation of (O3) are equivalent with the assumptions of case 1a in the present paper. Therefore, we should be able to compare (O3) with (7), with the understanding that $n \gg 1$. Using the Gaussian approximation (9) and Stirling's approximation (10), we can write (7) as:

$$M_n \cong \frac{1}{\sqrt{2\pi(n-1)}} \int_0^x \exp\left[-\frac{1}{2(n-1)}(y-n+1)^2\right] dy, \quad (n \gg 1) \qquad (19)$$

The integration limits in (19) are $y = 0$ and $y = x$. In order to compare (19) with (O3) we have to introduce a new integration variable ξ:

$$\xi = \frac{y}{c} - S \qquad (20)$$

such that the integration limits become $-S$ and D. Indeed, when $y = 0$, $\xi = -S$, and when $y = x$, $\xi = x/c - S = D$, according to eq. (1). In order to get (19) entirely in the form (O3) we replace the two constants n and c by two other constants M and h by making the substitutions:

$$n = 1 + 2h^2(M+S)^2 \qquad (21)$$

$$c = 2h^2(M+S) \qquad (22)$$

After substituting (20), (21) and (22) into (19) we get:

$$M_n \cong \frac{h}{\sqrt{\pi}} \int_{-s}^{D} \exp[-h^2(\xi - M)^2]\, d\xi \tag{23}$$

which is identical with (O3). The only non-essential differences between the two equations are that the integration variable is called x in (O3) and ξ in (23), and that the normalization constant C which is left undetermined in (O3) is written down explicitly in (23). The value of C is only equal to $h/\sqrt{\pi}$, though, if $n \gg 1$, but if this condition does not hold the derivations of both (O3) and (23) would be invalid anyway.

So far there seems to be complete agreement. The difficulties arise only through a particular choice of values for the as yet undetermined constants in (O3).

From eq. (21) follows that any choice of values for h, M and S determines the value of n. Since the whole derivation of (O3), and of (23) was based on the assumption of $n \gg 1$ we only have freedom to choose values for h, M and S as long as they are not in contradiction with the requirement $n \gg 1$. OPATOWSKI tries to make (O3) fit a straight line, and succeeds with the following values of the constants: $h = 2.5 \cdot 10^{-5}$ (roentgen)$^{-1}$, $M + S = 6642$ roentgen. After substitution of these values into (21) we obtain: $n = 1.055$, which certainly violates the requirement $n \gg 1$. Hence OPATOWSKI's conclusion, *i.e.*, that a many-event model can give a straight mutation frequency curve, is invalid. On the contrary, in trying to make (O3) fit a straight line it turns out that n is essentially 1, which is another proof that a straight line indicates a single event mechanism. That n does not turn out to be exactly 1 is a result of the fact that the Gaussian approximation (9) is rather poor for such a small n.

In a recent article by BOAG (1951), in which he points out the fallacy in OPATOWSKI's paper in a slightly different way, a sketch is made of the integrand $\exp[-h^2(x - M)^2]$ in (O3), using the previously mentioned values of the constants.* Due to the smallness of h this function falls off so slowly on both sides of the maximum, that at $x = -S$ the function still has practically maximum value. Thus it becomes clear why this function comes closest to an aproximation of a simple exponential e^{-x}, indicative for a single event mechanism.

SUMMARY

The question whether a linear relationship between mutation frequency and radiation dose is a strong indication that a mutation is caused by a single event has been investigated in several cases, and answered in the affirmative. The results of the calculations, supporting this conclusion, are shown in the

* The author is indebted to DR. BOAG for being able to see his manuscript before publication.

graphs of figure 1, in which the curves for multiple-event models are compared with the curve for a single-event model.

An exception has to be made for the possibility that some events may occur with much higher probability than any of the others, in which case a multiple-event model could give rise to a single-event type curve. However, this case is physically unlikely.

Finally it is shown that OPATOWSKI's curve, which is supposed to represent a many-event mechanism, actually represents a one-event mechanism, so that his conclusion that a many-event theory can give rise to a linear curve is unjustified.

ACKNOWLEDGMENTS

The author wishes to thank DR. CORNELIUS A. TOBIAS, DR. CURT STERN and DR. SEWALL WRIGHT for helpful suggestions and stimulating interest in this problem.

LITERATURE CITED

BOAG, J. W., 1951 On the interpretation of the dose-frequency curve in radiogenetics. Genetics **36**: 281–284.

OPATOWSKI, I., 1950 On the interpretation of the dose-frequency curve in radiogenetics. Genetics **35**: 56–59.

OPATOWSKI, I., and ALICE M. CHRISTIANSEN, 1950 On the single event hypothesis in radiogenetics. Bull. Math. Biophys. **12**: 19–26.

SPENCER, W. P., and C. STERN, 1948 Experiments to test the validity of the linear r-dose/mutation frequency relation in Drosophila at low dosage. Genetics **33**: 43–74.

W. P. A. FOR THE CITY OF NEW YORK, 1941 Tables of Prob. Functions, Vol. I.