*THE DETERMINATION OF L.D.50 AND ITS SAMPLING ERROR
IN BIO-ASSAY, III*

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated June 24, 1943

1. The discussion of this problem has been conducted on the hypothesis that the number n of animals used in each of the dilutions is the same. This restriction is convenient in that it leads to the possibility of preparing a table for the determination of the constants and of their standard errors when there are three dilutions.¹ Formulae can, however, be developed for the more general case where the numbers are not the same, but n_1, n_2, n_3 . The rule of maximum likelihood leads to the equations:

$$\frac{\partial L}{\partial \gamma} = -\alpha \Sigma (2s_i - n_i) + \alpha \Sigma n_i \tanh \alpha (x_i - \gamma) = 0, \quad (1)$$

$$\frac{\partial L}{\partial \alpha} = \Sigma (2s_i - n_i)(x_i - \gamma) - \Sigma n_i (x_i - \gamma) \tanh \alpha (x_i - \gamma) = 0. \quad (2)$$

Placing x_1, x_2, x_3 equal to $x - c, x, x + c$ and setting the derivatives equal to zero, introducing $X = \tanh \alpha (x - \gamma)$ and $C = \tanh \alpha c$ we have the simultaneous equations for X and C

$$C^2[n_2X^3 - AX^2 + (n_1 + n_3)X] + C(n_3 - n_1)(X^2 - 1) + A - NX = 0, \quad (3)$$

$$C^2[(n_3 - n_1)X - 2BX^2] + C(n_3 + n_1)(X^2 - 1) + 2B - (n_3 - n_1)X = 0, \quad (4)$$

where

$$A = 2(s_1 + s_2 + s_3) - N, \quad 2B = 2s_3 - 2s_1 - n_3 + n_1 \quad (5)$$

and $N = n_1 + n_2 + n_3$ is the total number of animals used. Here **A** and **B** which take the place of A and B in the earlier treatment are in terms of numbers, not of proportions, of the animals affected. The elimination of C gives a cubic equation for X only slightly more complicated than that found before, viz.,

$$\begin{aligned} Nn_2X^3 - [A(N + n_2) - 2B(n_3 - n_1)]X^2 \\ + [A^2 - 4B^2 + N(n_3 + n_1) - (n_3 - n_1)^2]X \\ - [A(n_3 + n_1) - 2B(n_3 - n_1)] = 0. \end{aligned}$$

Having solved this numerical equation for X , one may obtain C from the equation

$$C = \frac{-A(n_3 - n_1) + 2B(n_3 + n_1) + n_2(n_3 - n_1)X}{n_2(n_3 + n_1)X^2 - [A(n_3 + n_1) - 2B(n_3 - n_1)]X + 4n_1n_3},$$

and then

$$\alpha = \frac{1}{c} \tanh^{-1} C, \quad \gamma = x - \frac{1}{\alpha} \tanh^{-1} X,$$

where the functions $\tanh^{-1}$ may be replaced by their equivalents in logarithms.

The sampling errors may be obtained from

$$\frac{\partial^2 L}{\partial \gamma^2} = -\alpha^2 \Sigma n_i \operatorname{sech}^2 \alpha(x_i - \gamma) = -4\alpha^2 \Sigma n_i P_i Q_i,$$

$$\frac{\partial^2 L}{\partial \alpha^2} = -\Sigma n_i (x_i - \gamma)^2 \operatorname{sech}^2 \alpha(x_i - \gamma) = -4\Sigma n_i P_i Q_i (x_i - \gamma)^2,$$

$$\frac{\partial^2 L}{\partial \alpha \partial \gamma} = \alpha \Sigma n_i (x_i - \gamma) \operatorname{sech}^2 \alpha(x_i - \gamma) = 4\alpha \Sigma n_i P_i Q_i (x_i - \gamma),$$

when P_i are the fitted values, by

$$\sigma_\alpha^2 = -\frac{1}{H} \frac{\partial^2 L}{\partial \gamma^2}, \quad \sigma_\gamma^2 = -\frac{1}{H} \frac{\partial^2 L}{\partial \alpha^2},$$

where H is the Hessian of L with respect to γ and α . As a matter of fact these formulae are valid for any number of dilutions provided a solution for α and γ and consequently for P_i upon the curve has been found in any manner, presumably by successive approximation from the equations

$$\left. \frac{\partial L}{\partial \gamma} \right|_0 + \left. \frac{\partial^2 L}{\partial \gamma^2} \right|_0 \delta \gamma + \left. \frac{\partial^2 L}{\partial \alpha \partial \gamma} \right|_0 \delta \alpha = 0, \quad \left. \frac{\partial L}{\partial \alpha} \right|_0 + \left. \frac{\partial^2 L}{\partial \alpha \partial \gamma} \right|_0 \delta \gamma + \left. \frac{\partial^2 L}{\partial \alpha^2} \right|_0 \delta \alpha = 0.$$

2. In case one is working with a sufficient number of dilutions and in case the observed values P_i do not fluctuate too much from a straight line on growth paper one may get an approximate solution by graphical methods. For example, if we have the following logarithmic doses x and proportions of population affected²

x	1.755	1.784	1.811	1.837	1.861	1.884
n	28	27	30	31	30	29
s	9	14	23	29	29	29
P	0.329	0.519	0.767	0.936	0.967	1.000

we may plot the first 5 points ($P = 0$ or 1.0 cannot be plotted) and draw the line. The 50% end-point is read as at 1.777. The fitted growth curve being

$$P_f = 1/2 + 1/2 \tanh \alpha (x - 1.777), \quad \alpha = \frac{\tanh^{-1}(2P-1)}{x - 1.777}$$

permits the determination of α from any values P_f and x read from the line. At $x = 1.837$, $P_f = 0.935$. Hence $\alpha = 22.2$ and $P_f = 1/2 + 1/2 \tanh 22.2 (x - 1.777)$ is the line as determined graphically.

3. The values of P_f as fitted may be read³ from the line and the calculation form may be filled out.

x	1.755	1.784	1.811	1.837	1.861	1.884
P_f	0.280	0.587	0.810	0.935	0.975	0.992
Q_f	0.720	0.413	0.190	0.065	0.025	0.008
$P_f - P$	-0.049	0.068	0.043	-0.001	0.008	-0.008
$s_f - s$	-1.37	1.84	1.29	-0.03	0.24	-0.23
$x - \gamma$	-0.022	0.007	0.034	0.060	0.084	0.107
$(s_f - s)(x - \gamma)$	0.0301	0.0129	0.0439	-0.0018	0.0202	-0.0246
$4nP_fQ_f$	22.6	26.1	18.5	7.5	2.9	0.9
$4nP_fQ_f(x - \gamma)$	-0.50	0.18	0.63	0.45	0.25	0.10
$4nP_fQ_f(x - \gamma)^2$	0.011	0.001	0.021	0.027	0.021	0.011

Now,

$$\left. \frac{\partial L}{\partial \gamma} \right|_0 = 2\alpha \Sigma(s_f - s) = 77.3, \quad \left. \frac{\partial L}{\partial \alpha} \right|_0 = -2\Sigma(s_f - s)(x - \gamma) = -0.161,$$

$$\left. \frac{\partial^2 L}{\partial \gamma^2} \right|_0 = -38,700, \quad \left. \frac{\partial^2 L}{\partial \alpha^2} \right|_0 = -0.092, \quad \left. \frac{\partial^2 L}{\partial \alpha \partial \gamma} \right|_0 = 24.6.$$

Then $77.3 - 38,700\delta\gamma + 24.6\delta\alpha = 0$, $-0.161 + 24.6\delta\gamma - 0.092\delta\alpha = 0$ give as solutions $\delta\gamma = 0.00106$, $\delta\alpha = -1.47$ and $\gamma = 1.77806$, $\alpha = 20.73$ as a first arithmetic approximation subsequent to the graphical fit. If the graphical fit be considered an adequate approximation for computing σ_γ and σ_α we have for the Hessian $H = 2955$ and hence

$$\sigma_\alpha^2 = \frac{38,700}{2955} = 13.1, \quad \sigma_\gamma^2 = \frac{0.092}{2955} = 0.000031,$$

which gives $\sigma_\alpha = 3.6$, $\sigma_\gamma = 0.0056$ and we write⁴

$$\gamma = 1.7781 \pm 0.0056, \quad \alpha = 20.7 \pm 3.6.$$

In this example the logarithmic doses x did not advance by uniform amounts so that it is impossible satisfactorily to determine α and γ from the table which was computed on the basis of an equal advance in x and constant values of n ; but if one should ignore these points and take the 1st, 3rd and 5th observed P 's we should find $A = 1.126$, $B = 0.638$ giving " γ " = -0.63 and " α " = 1.00. With $c = 0.053$ we have $\gamma = -0.033$ and $\alpha = 19$. The 50% end-point would be at $1.811 - 0.033 = 1.778$. The values 1.778 and 19 in place of 1.777 and 22.2 determined graphically or 1.7781 and 20.7 obtained by the first arithmetic approximation are much the same thing in view of the standard errors estimated from the graphical solution. We could not expect to get the values of the standard deviations from the table because they would be estimated from only about one-half of the observations and furthermore the estimates depend on the actual points used on the growth curve as well as on the number of observations; in fact a calculation gives 1.778 ± 0.0080 and 19 ± 4.2 in place of 1.7781

± 0.0056 and 20.7 ± 3.6 . In this case, however, the use of the table and only three of the six observed proportions gives a fairly good answer.

4. There need be no restriction to three points or to identical values of n_i . If in (1) and (2) we write $x_i = x + ic$, assuming the logarithmic doses to proceed in a uniform manner (or the dilutions to be a geometric progression) we have for the determination of γ and α

$$\begin{aligned} -\alpha \Sigma(2s_i - n_i) + \alpha \Sigma n_i \tanh \alpha(x + ic - \gamma) &= 0, \\ \Sigma(2s_i - n_i)(x + ic - \gamma) - \Sigma n_i(x + ic - \gamma) \tanh \alpha(x + ic - \gamma) &= 0. \end{aligned}$$

If the first equation be multiplied by $(x - \gamma)/\alpha$ and added to the second, the latter becomes

$$\Sigma i(2s_i - n_i) - \Sigma i n_i \tanh \alpha(x + ic - \gamma) = 0.$$

The equations will be algebraic in $X = \tanh \alpha(x - \gamma)$ and $C = \tanh c$ but of too high an order to be readily used. It will, however, be noticed that the solution depends on the particular two constants

$$\mathbf{A} = \Sigma(2s_i - n_i), \quad 2\mathbf{B} = \Sigma i(2s_i - n_i).$$

For this reason, if some standard set-up involving more than three dilutions in geometric progression were to be adopted for wide-spread use it would be possible, and might be well, to compute a double entry table for⁵ γ and α and their standard deviations in terms of $\mathbf{A}$ and $\mathbf{B}$.

¹ These PROCEEDINGS, 29, 79-85, 114-120, 150-154, 207-212 (1943).

² For growth paper, see Wilson, E. B., these PROCEEDINGS, 11, 451-456 (1925). The data used in the illustration are taken from Bliss, C. I., p. 154 of *Ann. Applied Biol.*, 22, 134-167 (1935) or from Irwin, J. O., p. 17 of *Suppl. Jour. Roy. Stat. Soc.*, 4, 1-60 (1937). Those authors fit only the last six of the eight points. In the figure the limits of P at one standard deviation as determined from the formula on p. 209 of *Jour. Amer. Statis. Assoc.*, 22, 209-212 (1927) have been inserted and two lines have been drawn, one fitting (graphically) the last six and the other all eight points. It is clear that the graphical fit to the eight points is not bad; indeed if one actually fits the eight points arithmetically one finds $\gamma = 1.7714 \pm 0.0056$, $\alpha = 17.1 \pm 2.1$ with a χ^2 value of 3.35 which for six degrees of freedom gives $P = 0.76$ with a probability that because of the small numbers in some of the cells, the true value of P would be even larger. It is true that the deviations seem to be of such signs as to indicate curvature of the observed points to the fitted line, but in view of the small value of χ^2 and high value of P one might hesitate to draw that inference and to reject the first two points from this statistical evidence. The real reason for the rejection of these two points lies in the fact that in other corresponding sets of data the mortalities at low dosages seem also to be too high to fit the growth curve or probability integral curve. However, if the true form of the curve fitting such data were known, the values of the 50% end-point and of its standard error might be somewhat different from those found by fitting either of these curves to the last six points.

³ It might be better to calculate the values of P_i from the analytic expression even for the first arithmetic approximation to improve the value graphically determined, and calculation would certainly be necessary for a second arithmetic approximation.

⁴ If the arithmetic approximations were continued we should find $\gamma = 1.7787$, $\alpha = 20.53$. The value of χ^2 would be such as to give $P = 0.92$ for the fit to the six points, which would indicate that the variance of the observed from the fitted points was only 0.29 of the amount which on the average would be due to chance or that the variation of the observed from the fitted points would be only 0.54 of the chance amount.

⁶ For the actual computation of such a table one would not go back to the general and approximative equations which one would use to determine α and γ for given values of **A** and **B**. As the values of **A** and **B** are the same for the observed data and for the values fitted to them, one would start with assumed values of " γ " and " α " and determine **A** and **B** from the equations.

$$\mathbf{A} = \sum n_i \tanh " \alpha " (i - " \gamma "), \quad 2\mathbf{B} = \sum i n_i \tanh " \alpha " (i - " \gamma ").$$

This would give a table for **A** and **B** in terms of " γ " and " α " from which the table for " γ " and " α " could be obtained by inversion.