# REVIEW ARTICLE

## SIMPLIFIED MATHEMATICS FOR BIOASSAYS

By J. H. Gaddum, M.A., Sc.D., F.R.S.

*Professor of Pharmacology in the University of Edinburgh*

Several recent reviews and books deal with the application of mathematics to bioassays[1,2,3,4,5,6]. The object of the present paper is to review a few simple methods which can be used by those who have no deep knowledge of mathematics. These methods include those described in the British Pharmacopœia (1953), but the explanation is fuller and certain other methods are given. The object of a well-designed biological assay is to estimate the potency of an unknown or test preparation (U or T) by comparing its effect on living matter with that of a standard preparation (S). It is unfortunate that the letters U and T are used by different writers to mean the same thing. The letter U is used here although T is used in the B.P.

The present discussion is confined to the simplest case, where only one active substance is present, and the assay provides an estimate of its concentration in the unknown solution. In this case, potency and concentration are two words for the same thing and the experiment is called an analytical assay. When U and S contain different active substances it is also often possible to estimate their relative potency and such comparisons have played an important part in pharmacology since they have led to the discovery of new and better drugs. On the other hand, the results of such tests vary in ways which cannot be mathematically predicted and will not be considered here.

The result of a bioassay involves the assumption that living material behaves in a consistent way when exposed to drugs and the greater part of its error is generally due to the fact that this assumption is never quite true. The direct way of estimating the error is to repeat the assay a number of times and see how much the results vary among themselves. When the difference between duplicate estimates is consistently small it is reasonable to be content. The repetition of bioassays generally gives results which vary enough to raise doubts, and the assayist uses statistics of some kind with the double object of estimating the error and reducing it. One good way of estimating the error is to calculate the standard error of the result, when the experiment is repeated. When this is done directly, the estimate is obtained in arbitrary units, such as mg. per litre or units per ml., and does not mean much. It is generally better to calculate the standard error as a percentage of the result of the test, and this method is satisfactory when the percentage is less than 10. The percentage is however, often larger than this and may be greater than 100. It may be excusable in some cases to speak of an error of +150 per cent., but it would be absurd to speak of an error of −150 per cent. When the errors are large it is best to convert the result of each assay into a logarithm, and

to calculate the error from the variations occurring among these logarithms. This method of calculation is quite general and should be used in all cases, whether the error is large or small. It avoids the calculation of percentages and gives results in absolute units.

TABLE I

SIX INDEPENDENT ESTIMATES OF THE TOXICITY OF THE SAME PREPARATION OF NEOARSPHENAMINE AS PER CENT. OF THE STANDARD

| Toxicity per cent. | Log toxicity × 1000 | $d$ × 1000 | $d^2$ × $1000^2$ | |
|---|---|---|---|---|
| 101·3 | 2005 | 40·5 | 1640 | |
| 101·2 | 2005 | 40·5 | 1640 | |
| 77·3 | 1888 | 76·5 | 5852 | |
| 83·3 | 1921 | 43·5 | 1892 | |
| 93·0 | 1968 | 3·5 | 13 | |
| 99·9 | 2000 | 35·5 | 1260 | *Square roots* |
| | 6 \| 11787 | | 5 \| 12297 | |
| | 1964·5 | | 6 \| 2459·4 | 49·59 |
| | | | 409·9 | 20·25 |

Mean log potency = 1·9645 Potency = 92·15 per cent.

Standard deviation = $s = \sqrt{\frac{S(d^2)}{(n-1)}} = \frac{49{\cdot}59}{1000} = 0{\cdot}04959$

Standard error of mean = $\sqrt{s^2/n} = \frac{20{\cdot}25}{1000} = 0{\cdot}02025$

The calculations are illustrated in Table I which is based on some results obtained by Perry[7]. The toxicity of a preparation of neoarsphenamine was estimated 6 times and the results are shown on the left of the table. The differences among these results may be used to calculate the error. Each result is converted to a logarithm and this is multiplied by 1000 in order to avoid decimals. The mean (average) of these results is 1964·5. The third column shows the difference between each result and the mean. The fourth column shows the squares of these differences. When the figures in the fourth column are summed and divided by the number of degrees of freedom (n − 1), the result is an estimate of the variance of a single observation, which is another name for the square of the standard deviation. The square root of this (49·59) is an estimate of the standard deviation, and it must be divided by 1000 because the original figures were multiplied by 1000.

This standard deviation can be used to calculate the fiducial range. This depends on the arbitrary choice of the probability, P; if P = 0·95 then the results are expected to lie in the calculated fiducial range in 95 per cent. of experiments. The value of t is obtained from tables[1,2,8,9]. It depends on the number of degrees of freedom which have been used to estimate the standard deviation; if this number is small the estimate is unreliable and a comparatively large value of t must be used to allow for this. In the present case, the number of degrees of freedom is (n − 1) or 5; the corresponding value of t is 2·57. The fiducial range for a single estimate of log toxicity may be calculated as ±0·04959 × 2·57 or $\bar{1}{\cdot}8726$ to 0·1274. If 2 is added to these quantities, the corresponding antilogs

(74·6–134·1) give the range as a percentage of the estimate itself. For example, if the estimate is that the unknown drug is half as toxic as the standard the result is 50 (37·3–67·0) per cent. (P = 0·95).

The variance of the mean of the 6 results is $\frac{1}{6}$ of the variance of a single result and the corresponding standard error of the mean is 0·02025 (see Table I). The range for the mean is thus 1·9645 ± 0·02025 × 2·57 or 1·9124 — 2·0166. The table of antilogarithms gives the potency as 92·15 (81·7 — 103·9) per cent. (P = 0·95). These estimates of the error are somewhat larger than those calculated by Perry from the internal evidence of the assays.

When a calculating machine is available it is unnecessary to calculate the differences shown in the third column of Table I. In this case, the figures in the second column are squared and added and the result is 23167859. This figure is corrected by subtracting from it the product of the sum and average of these figures; thus 23167859 — 11787 × 1964·5 = 12297·5. This procedure depends on the fact that:

$$S(\bar{x} + d)^2 = S(\bar{x})^2 + 2\bar{x}\,S(d) + S(d^2) = n\bar{x}^2 + S(d^2) \text{ (since } S(d) = 0)$$

$$\therefore S(d^2) = S(\bar{x} + d)^2 - n\bar{x}^2$$

When two different workers or two different methods are compared, their results are unlikely to agree exactly and it may be asked whether the discrepancy between them can be accounted for by the errors of the tests. If the experiments are repeated often enough the results may be used to decide this point[10]. Each set of results is subjected to the processes outlined above. The variance of a single result is calculated by adding the two estimates of $S(d^2)$ and dividing by the total number of degrees of freedom.

$$V = \frac{S(d_1^2) + S(d_2)^2}{n_1 - 1 + n_2 - 1}$$

The variances of the two means are estimated as $V/n_1$ and $V/n_2$. The sum of these two quantities is an estimate of the variance of the difference of the means.

$$\therefore t = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{\frac{V}{n_1} + \frac{V}{n_2}}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the two means.

The significance of this quantity is determined by consulting a table of t using $(n_1 - 1 + n_2 - 1)$ degrees of freedom.

It is unusual to get even as many as 6 estimates of the potency of any one preparation of drug, but long series of duplicate estimates are sometimes available. Such results may give an accurate estimate of error, since each duplicate contributes one degree of freedom. Each result is converted to a logarithm and the difference between each pair of logarithms is calculated. These differences are squared and summed and divided by the total number of estimates (counting each duplicate as 2). The result is an estimate of the variance of the logarithm of a single

estimate, which can be used to calculate fiducial limits of a single estimate as in the example given above.

The above methods of estimating errors have two disadvantages. They are only reliable when experiments are repeated many times, and they involve the assumption that the error is always the same. It is better therefore, when possible, to estimate the error of each experiment from internal evidence by methods described below.

## Direct Assays

Preparations of digitalis or curare may be given by slow, intravenous injection until some definite effect is produced, so that the theshold dose is estimated in each animal. The best method of calculating the result and its error is like that described above. When an assay is complete, two sets of results will be available, one from experiments with the standard and the other from experiments with the unknown. Each result is converted into a logarithm. The difference between the means of these two sets of logarithms is used to calculate the result of the assay. Its variance is equal to the sum of the two variances.

## Assays Depending on Measured Effects

When the effect of the drug is measured the result depends on the relation between dose and effect. If the effect is plotted against log dose it is generally possible to draw a straight line which fits the results fairly

TABLE II

FORMULÆ FOR INTERPRETING PARALLEL LINE ASSAYS WITH 3 OR LESS DOSES OF EACH PREPARATION

| *Mean effects* | | *Effect differences due to* | | |
|---|---|---|---|---|
| Unknown | Standard | E Dose | F Preparation | G Slope difference |
| U | $S_1, S_2$ | $S_2-S_1$ | $U-\frac{1}{2}(S_1+S_2)$ | — |
| U | $S_1, S_2, S_3$ | $\frac{1}{2}(S_3-S_1)$ | $U-\frac{1}{3}(S_1+S_2+S_3)$ | — |
| $U_1, U_2$ | $S_1, S_2$ | $\frac{1}{2}(U_2-U_1+S_2-S_1)$ | $\frac{1}{2}(U_1+U_2-S_1-S_2)$ | $U_2-U_1-S_2+S_1$ |
| $U_1, U_2$ | $S_1, S_2, S_3$ | $\frac{1}{4}(U_2-U_1+2S_3-2S_1)$ | $\frac{1}{2}(U_1+U_2)-\frac{1}{3}(S_1+S_2+S_3)$ | $U_2-U_1-\frac{1}{2}(S_3-S_1)$ |
| $U_1, U_2, U_3$ | $S_1, S_2, S_3$ | $\frac{1}{3}(U_3-U_1+S_3-S_1)$ | $\frac{1}{3}(U_1+U_2+U_3-S_1-S_2-S_3)$ | $\frac{1}{2}(U_3-U_1-S_3+S_1)$ |

Log dose-interval = I  Slope = b = E/I
Log potency ratio (U/S) = M = F/b

| Design | Variances | | | |
|---|---|---|---|---|
| Doses | V | V(E) | V(F) = A | V(G) |
| 1 and 2 | $s^2/n$ or $1/wn$ | 2V | 3V/2 | — |
| 1 and 3 | | V/2 | 4V/3 | — |
| 2 and 2 | | V | V | 4V |
| 2 and 3 | | 2V/5 | 5V/6 | 5V/2 |
| 3 and 3 | | V/4 | 2V/3 | V |

With three doses, H (the index of curvature) = $S_1 + S_3 - 2S_2$ and V(H) = 6V.
$V(b) = V(E)/I^2$
Index of significance of $b = g = V(b)t^2/b^2$.
Fiducial limits of $M = \frac{M}{1-g} \pm \frac{t}{b(1-g)}\sqrt{A(1-g)+V(b)M^2}$

The potency ratio is the ratio $\frac{\text{estimated potency of U}}{\text{expected potency of U.}}$

The expected potency is that calculated on the assumption that the mean log dose of U is equivalent to the mean log dose of S.

well in a certain range of doses. In analytical assays the lines fitted to S and U should be parallel, and the result depends on the horizontal distance between them. This may be calculated in simple cases from the formulæ in Table II. The design using (2 and 2) doses (a 4-point assay[11]) is most generally used because it gives the smallest error for a given number of animals[8,12,16].

TABLE III

CALCULATIONS BASED ON DATA SHOWING THE FALL IN THE ASCORBIC ACID IN THE ADRENALS (MG./100 G. OF RAT) UNDER THE ACTION OF DIFFERENT DOSES OF ADRENOCORTICOTROPHIN

Measured effects (3 and 3) doses

| $U_1$ | $U_2$ | $U_3$ | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|---|---|
| 56 | 101 | 80 | 39 | 100 | 191 |
| 67 | 75 | 53 | 78 | 58 | 107 |
| 69 | 70 | 190 | −1 | 166 | 182 |
| 9 | 36 | 185 | 44 | 38 | 137 |
| 36 | — | 188 | 62 | — | 300 |
| Sum 237 | 282 | 696 | 222 | 362 | 917 |

| Mean effect | Sums | n | dn | Range | Range/dn |
|---|---|---|---|---|---|
| $U_1$ 47·4 | 257·1 | 5 | 2·33 | 60 | 25·8 |
| $U_2$ 70·5 | | 4 | 2·06 | 65 | 31·6 |
| $U_3$ 139·2 | | 5 | 2·33 | 137 | 58·8 |
| $S_1$ 44·4 | 318·3 | 5 | 2·33 | 79 | 33·9 |
| $S_2$ 90·5 | | 4 | 2·06 | 128 | 62·1 |
| $S_3$ 183·4 | | 5 | 2·33 | 193 | 82·8 |
| | diff. −61·2 | n = 28 / 4·667 | | sum | 295·0 |
| | | | | mean | 49·2 |

| | | |
|---|---|---|
| Degrees of freedom = S(n − 1) = 22 | $s^2/\bar{n}$ = 518 | V |
| $U_1 + U_2 + U_3 - (S_1 + S_2 + S_3)$ = −61·2 | ÷3 = −20·4 | F |
| $U_3 - U_1$ 91·8 } sum = 230·8 | ÷4 = 57·7 | E |
| $S_3 - S_1$ 139·0 } diff. = −47·2 | ÷2 = −23·6 | G |
| $U_3 + U_1 - 2U_2$ = 186·6 − 141·0 | = 45·6 | $H_1$ |
| $S_3 + S_1 - 2S_2$ = 227·8 − 181·0 | = 46·8 | $H_2$ |

*Result*

I = log dose interval = 0·301 = −0·106 = $\bar{1}$·894

M = F/b

b = E/I = 191·7

R = potency ratio = 0·783

| *Validity tests* | *Index* | *Variance* | *S.D.* | *t* | *P* |
|---|---|---|---|---|---|
| Parallelism | G = −23·6 | V = 518 | 22·8 | 1·04 | 0·3–0·4 |
| Curvature | $H_1$ = 45·6 | 6V = 3108 | 55·8 | 0·82 | 0·4–0·5 |
| | $H_2$ = 46·8 | | | 0·84 | 0·4–0·5 |

*Errors*

$\lambda = s/b = 49{\cdot}2/191{\cdot}7 = 0{\cdot}26$ $P = 0{\cdot}95$ $t = 2{\cdot}07$

$A = 2V/3 = 345{\cdot}3$ $V(b) = V/4I^2 = 1429$ $g = V(b)t^2/b^2 = 0{\cdot}167$

Fiducial limits $= \frac{M}{1-g} \pm \frac{t}{b(1-g)} \sqrt{A(1-g) + V(b)M^2} =$ 0·099 / $\bar{1}$·647

The potency of U is 78·3 (44–126) per cent. of the expected potency (P = 0·95).

| n | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| dn | 1·13 | 1·69 | 2·06 | 2·33 | 2·53 | 2·70 | 2·85 | 2·97 | 3·08 |

In routine work it is convenient to use standard forms for the calculations such as that shown in Table III for which a fairly elaborate example has been chosen in order to illustrate various special points.

In this assay the effects of three doses of standard adrenocorticotrophin were compared with those of 3 doses of an unknown preparation by the method of Sayers, Sayers and Woodbury[13]. Hypophysectomised rats

were anæsthetised and ascorbic acid was estimated in their adrenals before and after the injection of adrenocorticotrophin. The figures at the top show the fall ascorbic acid (mg. per 100 g./of rat) produced by the different doses of adrenocorticotrophin in 28 rats. In one case the result is negative because there was an apparent small rise.

The doses of each preparation were in the ratio 1:2:4 so that there was a constant ratio between neighbouring doses. It is important that this should be so, even when it involves rather awkward doses, such as 1, 1·2 and 1·44, otherwise the calculations become much more complicated.

The potency of U could be predicted with fair accuracy and doses were chosen so that the expected result was that each dose of U would have the same average effect as the corresponding dose of S. The result of the test (R) is the ratio of the estimated potency of U to the expected potency. It is convenient to calculate log R and this quantity is known as M, which is thus equal to the logarithm of the potency ratio.

The effects of each dose are added and the mean (average) effects are calculated by dividing by the number of effects in each group (n). These results are recorded in the second part of the table. The sum of the mean effects of all three doses of U is 257·1 and the corresponding sum for S is 318·3. When 318·3 is subtracted from 257·1 the result is −61·2 and this is divided by 3 to give an estimate of F, which is the mean difference between the effects of U and S, and is negative because U has less mean effect than S and is therefore estimated to be less active.

The meaning of this estimate of F (−20·4) in terms of potency is calculated from the results obtained with different doses of each preparation. The quantity I is the logarithm of the dose-ratio or the difference between the logarithms of two neighbouring doses and in this case is equal to log 2 or 0·301. The quantity E is an estimate of the difference of effect corresponding to I, so that the slope of the curve (b) may be estimated as E/I. The results with U show that the difference between the effects of the largest and smallest doses ($U_3 - U_1$) was 91·8 and this corresponds to two log dose intervals (2I) so that the slope of the curve was ($U_3 - U_1$)/2I. The middle dose contributes nothing to the estimate of slope and its effect can be neglected. The slope of the standard curve was ($S_3 - S_1$)/2I and the average slope is obtained by adding these figures and dividing by 2.

$$b = \frac{U_3 - U_1 + S_3 - S_1}{4I} = \frac{E}{I} = 191{\cdot}7.$$

A difference of 191·7 on the scale of effects is thus estimated to correspond to a difference of 1 on the scale of log dose. A difference of −20·4 in effects (F) therefore corresponds to a difference of $\frac{-20{\cdot}4}{191{\cdot}7}$ (= F/b) in the scale of log dose and this is an estimate of M — the logarithm of the potency ratio. The result is M = −0·106 = $\bar{1}$·894 which is the logarithm of 0·783. This is the result of the test. The potency of U is estimated to be 0·783 times, or 78·3 per cent. of that of S.

The rest of the calculations provide tests of validity and an estimate of

error. These depend on an estimate of the variation of the effects among the animals receiving the same dose. The standard deviation of a single effect (s) could be estimated by calculating from the figures at the top the deviations from the mean within each dose (d) and using the formula $s^2 = S(d^2)/S(n - 1)$ and this gives the result that $s = 51{\cdot}1$. Table III shows a quicker way of obtaining much the same result. The range of effects in each group is calculated by subtracting the smallest effect from the largest effect; thus, for $U_1$, $69 - 9 = 60$. The quantity $d_n$ is the ratio of expected range to the standard deviation, and this depends on the number of observations (n). The value of $d_n$ for each group is copied from the figures given at the bottom of Table III. The numbers obtained by dividing the range by $d_n$ are all estimates of s and their mean (49·2) is the estimate used in the later calculations. This figure agrees reasonably well with the figure obtained by the best possible method in which squares of the deviations are summed (51·1). This may be expected to be so when $n < 10$. When n is constant it is unnecessary to calculate each value of the ratio range/$d_n$ separately since the same result may be achieved by calculating the mean range and dividing this by $d_n$.

The number of degrees of freedom contributed by each group of results is $(n - 1)$ and the total number on which the estimate of s is based is therefore $S(n - 1)$ or 22.

The figures given here might be taken to indicate that s increases as the dose increases. For the present purpose, it is assumed that the differences between the estimates obtained with different doses are due to chance, and that the standard deviation of the effect does not really depend on the dose. The results of other experiments seem to show that this assumption is probably correct.

The variance (square of the standard deviation) of the mean of n effects (V) is estimated as $s^2/n$. In this experiment n was not constant and ideally, the calculations should be made much more complicated in order to allow for this. In the method recommended here, V is assumed to be constant and equal to $s^2/\bar{n}$, where n is the mean number of effects per dose.

The quantities G and H provide tests of validity. G is a measure of the difference between the two slopes; in fact, G/I is equal to this difference. Since $G = \dfrac{(U_3 - U_1) - (S_3 - S_1)}{2}$, and since the variance of each of the quantities in the top line is V, the variance of G is $\dfrac{4V}{2^2} = V = 518$. The standard deviation of G is therefore $\sqrt{518}$ or 22·8. The ratio of G to its standard deviation is $23{\cdot}6/22{\cdot}8 = 1{\cdot}04 = t$. The significance of this value is tested by consulting a table of t with 22 degrees of freedom, since the estimate of V is based on 22 degrees of freedom. In this case, the table shows that P lies between 0·3 and 0·4. The curves are, of course, not exactly parallel, but the difference of their slopes is no larger than might reasonably have been expected. If G had been significant then the calculations would be shown to be based on false assumptions. Such a result is an indication of a qualitative difference between U and S.

The quantity $H_1$ is an index of the curvature of the line for U. If this line was straight the point midway between $U_3$ and $U_1$ would correspond to an effect $(U_3 + U_1)/2$. The difference between this and $U_2$ is $(U_3 + U_1)/2 - U_2$ which is equal to $H_1/2$. $H_1$ is thus equal to twice the vertical distance between the point corresponding to the effect of the middle dose and the line joining the points corresponding to the other two doses. If $H_1 = 0$ the line is straight. If $H_1$ does not differ significantly from 0 the line may be straight. The variance of $H_1$ is equal to the sum of the variances of $U_3$, $U_1$, and $2U_2$ which is $V + V + 4V = 6V$. The significance of the curvature is determined by calculating t = the ratio of $H_1$ to its standard deviation $= H_1/\sqrt{6V} = 0{\cdot}82$ which is not significant for 22 degrees of freedom. The quantity $H_2$ is similarly an index of the curvature of the other curve. When these quantities are positive the midpoint lies below the line joining the other two and *vice versa*. When $H_1$ and $H_2$ have the same sign their combined evidence may prove curvature. This may be tested by adding them (92·4) and comparing this sum with its estimated standard deviation $(\sqrt{12V})$. In this case, $t = 1{\cdot}17$ and $P = 0{\cdot}2 - 0{\cdot}3$ which is not significant. When $H_1$ and $H_2$ have different signs their difference should similarly be compared with its standard deviation $(\sqrt{12V})$.

The error of the test may be expressed in the form of the fiducial limits corresponding to a given probability (P). In the present case, $P = 0{\cdot}95$ and the actual potency may be expected to lie within the calculated fiducial limits in 95 per cent. of assays. The result of the assay has been estimated from the formula $M = E/FI$ and the expression for the fiducial limits is complicated because both E and F are subject to error. This expression is given in Tables II and III in its general form. The justification for using this expression is rather complicated and will not be given here. It is discussed in a paper by Irwin[14]. In order to use this formula it is necessary first to calculate A, V(b) and g. A is the variance of F and since F is calculated by adding 6 quantities with variance V and dividing by 3, $A = 6V/3^2 = 2V/3$. V(b) is the variance of b and since b is calculated by adding 4 quantities with variance V and dividing by 4 and $I^2$, $V(b) = 4V/16I^2 = V/4I^2$. The quantity g is the index of significance of b and is defined as $V(b)t^2/b^2$. If $g > 1$, then b is not significantly different from zero and the fiducial limits are $\pm$ infinity. If $g < 0{\cdot}1$ it can be neglected and the formula for the fiducial limits becomes simpler.

The calculation of the fiducial limits is shown in Table IV. These calculations show that it is reasonable to conclude that the potency of U lies between 44 and 126 per cent. of the potency expected when the test was designed. Such conclusions are likely to be correct in 95 per cent. of cases.

## Twin Crossover Tests

This design was introduced by Smith, Marks, Fieller and Broom[15] for assays of insulin on rabbits. It is also suitable for use in other cases where the effect of each drug may be measured on each animal. The

same group of rabbits is used for two separate (2 and 2) dose assays, and the animals are crossed over, so that the animals which receive the unknown in the first assay receive the standard in the second, and those which receive a small dose in the first assay receive a large dose in the second, and *vice versa*. The calculations are based entirely on the difference between the two effects observed on each rabbit (y) so that

TABLE IV

CALCULATION OF FIDUCIAL LIMITS FOR TABLE III

| | | | | |
|---|---|---|---|---|
| $A(1-g)$ | $345 \times 0{\cdot}833$ | | | $= 287$ |
| $V(b)\,M^2$ | $1429 \times (0{\cdot}106)^2$ | | | $= 16$ |
| | | | Sum | $= 303$ |
| | | | Square root | $= 17{\cdot}41$ |
| $\times \frac{t}{b(1-g)}$ | $\times \frac{2{\cdot}07}{191{\cdot}7 \times 0{\cdot}833}$ | $(= 0{\cdot}01298)$ | | $= 0{\cdot}226$ |
| $\frac{M}{(1-g)}$ | $\frac{-0{\cdot}106}{0{\cdot}833} = -0{\cdot}127$ | | | $= \bar{1}{\cdot}873$ |
| | Upper limit | $1{\cdot}256$ | Sum | $= 0{\cdot}099$ |
| | Lower limit | $0{\cdot}4436$ | Difference | $= \bar{1}{\cdot}647$ |

TABLE V

ARRANGEMENT OF DOSES IN A TWIN CROSS-OVER TEST

| Group of animals | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| First assay | $S_1$ | $S_2$ | $U_1$ | $U_2$ |
| Second assay | $U_2$ | $U_1$ | $S_2$ | $S_1$ |
| Mean response difference $(U - S)$ | $y_1$ | $y_2$ | $y_3$ | $y_4$ |

differences between rabbits do not affect the result. (See Table V.) All the values of y within each group of animals would be the same, if the animals did not vary. The variance within these groups (y) is calculated from the formula $V(y) = S(d^2)/S(n-1)$. The result and its error can be calculated from Table II taking

$$E = \tfrac{1}{4}(y_1 - y_2 - y_3 + y_4)$$
$$F = \tfrac{1}{4}(y_1 + y_2 + y_3 + y_4)$$
$$V(E) = V(F) = V(y)/4n$$

The reasons for using these formulæ for E and F may be seen by substituting $(U_2 - S_1)$ for $y_1$, etc., in them. The variance of E (or F) is equal to the sum of the variances of $y_1$, $y_2$, $y_3$ and $y_4$ divided by $4^2 = \frac{V(y_1).}{4}$

The variance of $y_1$ is $V(y)/n$, since $y_1$ is the mean of n values.

## THE METHOD OF THE CONSTANT STANDARD

Vos[16] has described a convenient method of calculating the result and error of an assay with an isolated organ such as a uterus in a bath. In this method a constant dose of the standard alternates with varying doses of the unknown. Effect differences are calculated by subtracting from each effect of the unknown the mean of the two neighbouring effects of the standard. The calculations are designed to estimate the dose of the unknown for which the effect difference is zero. This is estimated

to be equivalent to the standard dose of the standard. Table II may be used to calculate the results of such experiments and their error.

In this case d = the difference of each effect difference from the mean of the dose group to which it belongs.

$s^2$ = the variance of the effect difference = $\frac{S(d^2)}{S(n-1)}$

This can also be calculated from ranges.

F = the mean of all the effect differences.

A = V (F) = $s^2/N$, where N = the total number of effect differences. If $U_1U_2$ etc., are the mean effect differences

For 2 doses $E = U_2 - U_1$ $V(E) = 2V$

3 doses $E = \frac{1}{2}(U_3 - U_1)$ $V(E) = V/2$

4 doses $E = \frac{1}{10}(3U_4 + U_3 - U_2 - 3U_1)$ $V(E) = V/5$

## Assays Depending on Quantal Effects

The effect of the drug is sometimes measured in terms of the percentage of the animals which give a definite response of some kind. If this percentage is converted into a probit by means of suitable tables[1,8,9] and if these probits are plotted against log dose, it is generally possible to fit the results fairly well with straight lines, or to calculate the potency and slope by methods like those described above. The best possible answer can only be obtained by means of complicated calculations in which each

TABLE VI

Calculations for an assay of strophanthus tincture by its lethal action in frogs[17]

Quantal effects (2 and 2) doses

| | Effects | n | Per cent. | Probit | Sums | w | wn |
|---|---|---|---|---|---|---|---|
| $U_1$ | 7 | 20 | 35 | 4·61 | 11·25 | 0·60 | 12 |
| $U_2$ | 19 | 20 | 95 | 6·64 | | 0·22 | 4·4 |
| $S_1$ | 2 | 20 | 10 | 3·72 | 9·56 | 0·34 | 6·8 |
| $S_2$ | 16 | 20 | 80 | 5·84 | | 0·49 | 9·8 |
| | | | | | diff. 1·69 | | Sum 33·0 |
| | | | | | | | $\overline{wn}$ = 8·25 |

Variance of each probit = $1/\overline{wn}$ = 0·1212 V

$U_1 + U_2 - (S_1 + S_2)$ = 1·69 ÷2 = 0·845 F

$U_2 - U_1$ 2·03 sum 4·15 ÷2 = 2·075 E

$S_2 - S_1$ 2·12 diff. 0·09 = 0·09 G

$G/2\sqrt{V}$ = $0·09/2\sqrt{0·121}$ = 0·13 t(G)

>0·8 P(G)

*Result*

I = log dose interval = 0·176 E/I = 11·8 b

F/b = 0·0717 M

Potency ratio = 1·18 R

*Errors*

$\lambda = 1/b = 0·85$ V = 0·1212 A

P = 0·95 $V/I^2$ = 3·91 V(b)

t = 1·96 $V(b)t^2/b^2$ = 0·108 g

Log fiducial limits = $\frac{M}{1-g} \pm \frac{t}{b(1-g)}\sqrt{A(1-g) + V(b)M^2}$ = 0·1469 / 0·0139

The potency of U is 118 (103–140) per cent. of the expected potency (P = 0·95).

observation receives an appropriate weight, and the final result depends on a series of successive approximations. A result which is accurate enough for most purposes can be obtained by assuming that all the weights are equal to the average weight. The calculations are then exactly the same as those described above. They will be illustrated by two examples. The first is a simple (2 and 2) dose assay, and the second is a more complicated example which shows some of the difficulties which may arise.

The calculations for the first example are shown in Table VI. In this assay[5,17] two tinctures of strophanthus (U and S) were diluted with saline solution and injected into frogs. The result depended on counting the number of deaths 24 hours later. Two dilutions of each tincture were used and these were chosen so that each tincture was expected to produce about the same effect.

TABLE VII

CALCULATIONS BASED ON DATA[3] SHOWING THE NUMBER OF RATS WHICH BECAME FERTILE UNDER THE ACTION OF VITAMIN E. WEIGHTS ASSUMED CONSTANT (METHOD 4)

Quantal effects (3 and 3) doses

| | Effects | n | Per cent. | Probit | Sums | w | wn |
|---|---|---|---|---|---|---|---|
| $U_1$ | 0 | 10 | 0 | 2·60 | | — | 1·81 |
| $U_2$ | 2 | 12 | 16·7 | 4·03 | 11·88 | 0·448 | 5·38 |
| $U_3$ | 6 | 10 | 60 | 5·25 | | 0·622 | 6·22 |
| $S_1$ | 2 | 8 | 25 | 4·33 | | 0·538 | 4·30 |
| $S_2$ | 4 | 8 | 50 | 5·00 | 16 66 | 0·637 | 5·10 |
| $S_3$ | 8 | 8 | 100 | 7·33 | | — | 1·63 |
| | | | | | diff. −4·78 | | Sum 24·44 |

$V = \frac{1}{\text{Mean wn}} = 0{\cdot}246$ diff. ÷ 3 = −1·59 F

$U_3 - U_1$ 2·65 ⎫ Sum 5·65
$S_3 - S_1$ 3·00 ⎭ Diff. −0·35

÷ 4 = 1·41 E
÷ 2 = −0·175 G

$U_3 + U_1 - 2U_2 = 7{\cdot}85 - 8{\cdot}06$ = −0·21 $H_1$
$S_3 + S_1 - 2S_2 = 11{\cdot}66 - 10$ = 1·66 $H_2$

*Result*

I = 0·176 b = E/I = 8·0 M = F/b = −0·199 = $\bar{1}$·801
R = potency ratio = Antilog M = 0·632

| *Validity tests* | *Index* | *Variance* | *SD* | *t* | *P* |
|---|---|---|---|---|---|
| Parallelism | G = −0·175 | V = 0·246 | 0·496 | 0·35 | 0·7–0·8 |
| Curvature | $H_1$ = −0·21 | 6V = 1·476 | 1·215 | 0·17 | 0·8–0·9 |
| | $H_2$ = 1·66 | | | 1·36 | 0·1–0·2 |

*Errors*

λ = 1/b = 0·125 P = 0·95 t = 1·96
A = 2V/3 = 0·164 V(b) = V/4I² = 1·98 g = V(b)t²/b² = 0·119

Fiducial limits of $M = \frac{M}{1-g} \pm \frac{t}{b(1-g)} \sqrt{A(1-g) + V(b)M^2} = \begin{matrix} \bar{1}{\cdot}9055 \\ \bar{1}{\cdot}6427 \end{matrix}$

The potency of U is 63·2 (43·9 − 80·4) per cent. of the expected potency (P = 0·95).

The small dose of the unknown tincture ($U_1$) killed 7 frogs out of 20 or 35 per cent. It was found, by consulting the appropriate tables[1,2,5,8,9], that the corresponding probit was 4·61 and the weight factor (w) 0·60. The weight of this observation (wn) is obtained by multiplying w by the total number of animals in the group (20). The mean weight (wn) is calculated by adding the separate weights and dividing by their number (4). The variance of a single observation (V) is the reciprocal of the mean weight. The rest of the calculations are similar to those already discussed (Table III) but simpler because only (2 and 2) doses are used.

Since the variance is calculated theoretically, the number of degrees of freedom used with t is infinite.

The second example with quantal effects is taken from an experiment recorded by Bliss[3] (p. 549) in which vitamin E was assayed by counting the number of rats which became fertile under its influence. The results of this experiment are shown in the upper part of Table VII. It will be seen that 3 doses of each preparation were used with a constant dose ratio, but that the smallest dose of U had no effect and the largest dose of S had 100 per cent. effect. These two results cannot be used in the same way as the other results, since the corresponding probit is infinite and its weight zero. One way of dealing with such results is to neglect them and calculate the answer as if the experiment had originally been designed as a (2 and 2) dose assay. The results obtained in this way are shown in the top line of Table VIII. The best solution of this problem is probably that given by the method of maximum likelihood, using successive approximations. Table VIII shows the results of Bliss' calculations using this method. The third approximation is practically the same as the second approximation and must be very close to the best possible answer. Second and third approximations for V(b), g and the range are not given here because they were not given by Bliss. It would not be worth calculating them very exactly, because they inevitably involve the approximation that probits are normally distributed.

TABLE VIII

RESULTS OBTAINED BY DIFFERENT METHODS FROM THE DATA IN TABLE VII

| Method of calculation | b | V(b) | g | Potency of U (percent. of S) with fiducial limits (P = 0·95) |
|---|---|---|---|---|
| 1. Neglecting 0 and 100 per cent. .. | 5·37 | 6·14 | 0·82 | 66 (26·6 — 148) |
| 2. Bliss[3] 1st approx. .. .. | 7·0 | 2·53 | 0·2 | 65 (44·5 — 84) |
| 2nd approx. .. .. | 7·345 | — | — | 64·5 |
| 3rd approx. .. .. | 7·364 | — | — | 64·36 |
| 3. Gaddum[17] .. .. .. | 7·35 | 2·72 | 0·19 | 64·22 (40·1 — 83·5) |
| 4. Same with constant weights .. | 8·0 | 1·98 | 0·119 | 63·2 (43·9 — 80·4) |

It will be seen that the simple solution in the top line gives a good estimate of R in this case, but it underestimates b. The neglected effects show that the lines are steeper than might otherwise have been thought. The most important consequence of this is the effect on g, which is 0·82 in the top line instead of 0·2. This means that the variance of b is very large compared with b itself and has the effect of greatly increasing the fiducial range.

The next estimate (3) in Table VIII was obtained by a method recommended some time ago[17], but not much used. In this method, a percentage of 0 or 100 is represented in the calculations by a probit and weight which depend only on the number of animals in the group. In the original description of this method, these values were obtained from a graph. It may perhaps be more convenient to have these figures in the form given in Table IX. Estimate 3 was obtained by calculating the formula of the regression line directly, using this table when the percentage was 0 or 100. For other percentages, the weight must be that appropriate to the observed

percentage and not that calculated from a preliminary fitting of the line, as in the method for obtaining maximum likelihood. It will be seen that the results obtained by method 3 agreed well, in this case, with the results obtained by Bliss[3]. Similar comparisons have been made with the results obtained by Armitage and Allen[18] using various different methods of calculation. The results given by method 3 agreed well in every case with the maximum likelihood solution and less well with the various other solutions calculated by these authors. The results of this method are

TABLE IX

VALUES OF PROBIT AND WEIGHT USED WHEN THE PERCENTAGE IS 0 OR 100

| n | Probits | | Weight |
|---|---|---|---|
| | 0 per cent. | 100 per cent. | wn |
| 1 | 3·36 | 6·64 | 0·53 |
| 2 | 3·13 | 6·87 | 0·82 |
| 3 | 2·99 | 7·01 | 1·02 |
| 4 | 2·90 | 7·10 | 1·19 |
| 5 | 2·82 | 7·18 | 1·32 |
| 6 | 2·76 | 7·24 | 1·44 |
| 7 | 2·71 | 7·29 | 1·54 |
| 8 | 2·67 | 7·33 | 1·63 |
| 9 | 2·63 | 7·37 | 1·72 |
| 10 | 2·60 | 7·40 | 1·81 |
| 12 | 2·54 | 7·46 | 1·93 |
| 15 | 2·47 | 7·53 | 2·10 |
| 18 | 2·41 | 7·59 | 2·24 |
| 20 | 2·38 | 7·62 | 2·32 |
| 24 | 2·32 | 7·68 | 2·46 |
| 25 | 2·31 | 7·69 | 2·50 |
| 30 | 2·26 | 7·74 | 2·66 |
| 40 | 2·17 | 7·83 | 2·90 |
| 50 | 2·10 | 7·90 | 3·10 |
| 60 | 2·05 | 7·95 | 3·25 |
| 70 | 2·01 | 7·99 | 3·39 |
| 80 | 1·97 | 8·03 | 3·52 |
| 90 | 1·93 | 8·07 | 3·64 |
| 100 | 1·90 | 8·10 | 3·76 |

These values were calculated as follows. Let $f_1$ and $f_2$ be the ordinates of the normal curve corresponding to 1 and 2 standard deviations as a proportion of the maximum ordinate (approx. 0·607 and 0·135 respectively). Calculate $p_1$ and $p_2$, where $p_1 n = f_1$ and $p_2 n = f_2$ and n = the number of animals observed. Consult tables to find the probits ($y_1$ and $y_2$) corresponding to $p_1$ and $p_2$. Then the probit given in Table VII for the case where all die is $(2y_1 - y_2)$ and the weight is $1/(y_1 - y_2)^2$.

accurate enough for practical purposes and the calculations are only done once (apart from essential checking). If it is essential to be certain that maximum likelihood has been achieved this method of calculation provides a very good first approximation. The results do not depend on a preliminary graphical fitting of the line, though it is always best to plot the results in order to be certain that the calculations have given a reasonable answer. It may be found when this is done that the probit given in Table IX lies nearer to 5 than might have been expected from the regression lines. Such results merely confirm that extreme doses produce extreme effects. They contribute no new information and should be neglected. They are unsuitable for calculations depending on Table IX.

In method 4 (Table VIII) the percentages of 0 and 100 are treated as in method 3 using Table IX, but the calculations are simplified (as in Table VI) by assuming the weight constant. This device is already widely used in connection with 4-point assays. The calculations are shown in Table VII. The results agree well with the results obtained by much more

elaborate calculations and this method can be recommended for general use, at any rate as a first approximation.

The values of H in Table VII indicate the shape of the curves, and provide another method of excluding extreme doses of the kind discussed above. When a percentage of 0 is included, H should be negative as it is with the unknown in Table VII; when a percentage of 100 is included H should be positive as it is with the standard. If H has the wrong sign, and is positive when it should be negative or negative when it should be positive, methods depending on Table IX cannot be used. In such cases, it may be possible to use the results by neglecting the 0 or 100 altogether and calculating the result as a (2 and 3) dose assay.

## REFERENCES


1. Emmens, *Principles of Biological Assay*, Chapman and Hall, London, 1948.
2. Burn, Finney and Goodwin, *Biological Standardization*, 2nd edition. Oxford University Press, London, 1950,
3. Bliss, *Vitamin Methods*, 1951, **2**, 445.
4. Emmens, *Hormones—A Survey of Their Properties and Uses.* Chap. V. Standardization. The Pharmaceutical Press, London, 1951.
5. Finney, *Statistical Method and Biological Assay.* Charles Griffin, London, 1952.
6. Gaddum, *Pharmacol. Rev.*, 1953, **5**, 87.
7. Perry, *Spec. Rep. Ser. med. Res. Coun. Lond.*, 1950, **270**, 1.
8. *British Pharmacopœia*, 1953.
9. Fisher and Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, Edinburgh, 1953.
10. Gaddum, *Biochem. J.*, 1931, **25**, 1113.
11. Bacharach, Coates and Middleton, *Biochem. J.*, 1942, **36**, 407.
12. Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, 1939, **12**, 82.
13. Sayers, Sayers and Woodbury, *Endocrinology*, 1948, **42**, 379.
14. Irwin, *J. Hyg.*, 1943, **43**, 121.
15. Smith, Marks, Fieller and Broom, *Quart. J. Pharm. Pharmacol.*, 1944, **17**, 108.
16. Vos, *J. Amer. pharm. Ass.*, 1943, **32**, 138.
17. Gaddum, *Spec. Rep. Ser. med. Res. Coun. Lond.*, 1933, **183**, 1.
18. Armitage and Allen, *J. Hyg.*, 1950, **48**, 298.