BIOASSAYS AND MATHEMATICS

J. H. GADDUM

Department of Materia Medica and Pharmacology, University of Edinburgh

I. NOTATION

S ( )	The sum of all values of the expression in the bracket.
X	Dose.
x	Dose metameter (often $\log_{10} X$). (92).
I	Log-dose interval. (40).
t	(1) Time (53).
	(2) Students t. A quantity divided by its estimated standard deviation.

Y	Effect as first measured.
y	Effect metameter. Function of Y used in calculations (*e.g.*, probit). (92).
x_s y_s x_u y_u	Observed values with standard and unknown.
S_1, S_2, U_1, U_2 etc.	Observed mean values of y_s or y_n for different doses. (40).
Z	Observed mean value of y for zero dose.
$\bar{x}$, $\bar{y}$	Weighted mean of all observed values with one preparation. (92).
$b = \frac{dy}{dx}$	Slope of the line connecting y and x.
$a = \bar{y} - b\bar{x}$	Value of y when $x = 0$. Zero ordinate.
3/10	Three quantal effects out of 10 trials.
p	Probability of a quantal effect.
n	Number of observations in one dose group. (92).
N	Total number of observations in one experiment. (92).
s	Estimate of standard deviation of y. (39).
w	Weight factor for a probit (from tables). (83).
$W = n/s^2$ or wn	Weight of mean response to one dose.
$V = 1/W$	Variance of mean response to one dose.
I.E.D.	Individual effective dose. Tolerance.
λ	Standard deviation of I.E.D., estimated from s/b (measured effects) or $1/b$ (quantal effects). (91).
L	$1/\lambda$. (204)
ED50	Dose causing 50 per cent of quantal effects. (181, 96).
m	Log ED50. (92).
R	Potency ratio $= \frac{\text{Estimated potency of unknown}}{\text{Assumed potency}}$ (203).
M	Log R. (92).
M', R'	Most likely values of M and R.
$s_{\bar{y}}$, s_b, s_M, etc.	Estimates of standard deviation of $\bar{y}$, b, M, etc.
$V(\bar{y})$, $V(b)$, $V(M)$	$= s_{\bar{y}}^2$, s_b^2, s_M^2 Estimates of variance.
g	$V(b)t^2/b^2$. (78).
A	Estimated variance of difference of mean effects. $(V(\bar{y}_s) + V(\bar{y}_u))$ (124).
B^2	Variance due to slope. (40).
C	(1) Concentration. (53). (2) $1/(1-g)$ (71). (3) $\sqrt{1/(1-g)}$ (Bliss) (36).
D^2	Variance due to preparation. (40).
E	Dose difference } See Table 5.
F	Preparation difference } See Table 5.
G	Slope difference } See Table 5.

II. INTRODUCTION

The mathematical methods used in bioassays have become so complicated that few pharmacologists can keep track of them. The professional mathematicians have developed new methods which attract some people because they give greater accuracy with fewer animals, but repel others because of the mental effort needed to understand them and because of the labour involved in using some of them. Few pharmacologists are willing to follow the mathematicians through all the complexities which lead to their ideal solution, and in practice it is usually desirable to make a compromise between the conflicting claims of precision and simplicity, but it is not always easy to decide just where to draw the line.

The mathematical techniques are mostly described in standard textbooks of statistics (87, 175) and have been fully discussed in three monographs (46, 68, 84A) devoted to bioassays in general. Special monographs nominally devoted to hormones (69), vitamins (36, 57), and probits (83) also include general discussions. These books may be consulted for detailed descriptions of the methods recommended by the leading authorities. Comprehensive bibliographies for 1940–1947 have also been published (67, 3). The object of the present review is to provide a more general guide to the literature of the subject, and a discussion of some of the principles involved. This literature was fully reviewed in 1943 by Bliss and Cattell (38). Many of the pharmacological techniques are described in detail in the book by Burn, Finney and Goodwin (46).

The words "bioassay" and "biological standardization" may be applied to any experiment in which the potency of a drug is measured by its effect on living organisms or tissues. In the sim lest case, the result gives the concentration of a single known substance in a solution. In this case there can be no significant difference between the results obtained by different method, but when the potent substances are unknown or variable, the result generally varies when the technique varies, or even when it is kept as constant as possible. It is well known for example that if vitamin D_2 is used as a standard, fish-liver oils appear more potent in assays on chickens than in assays on rats. This kind of discrepancy is so common that agreemen between parallel quantitative tests is sometimes taken as evidence that the active substance has been correctly identified (50, 101).

The methods developed in connection with these "analytical" assays have also been applied to "comparative" assays in which measurements are made of the potency of new drugs, or mixtures of variable composition (82A). It is important to realize that such estimates of potency are likely to depend on the details of the technique. This is true even when two simple similar substances like acetylcholine and propionylcholine are compared with one another (50) and it is therefore not surprising that serious difficulties arise in assays of digitalis or of bacterial toxins which may contain variable mixtures of active substances known and unknown.

Comparative assays have no satisfactory logical basis (130, 148) and their

only justification lies in the fact that the greatest contributions of pharmacology to medicine have been founded upon them. Under the name of screening tests they have led to the synthesis of many potent drugs, but their results have not the same general validity as the results of analytical assays. Special tests may sometimes show that an estimate of relative potency depends on the dose used. Such tests provide an important check on the validity of the assumptions used in the calculations, but a failure to demonstrate invalidity does not prove much. If other factors (species, type of effect, etc.) were varied the results of most comparative assays could be shown to be invalid.

The best that can be expected in a comparative assay is a method which is found by experience to give a reliable prediction of what will happen when a drug is used in a particular way. This is most likely to be achieved if the conditions of the assay are similar to the conditions found in practice. Most of the methods of assay now in use depend on experiments on animals, but assays on man are more likely to give a reliable prediction of effects on man and similar methods have recently been used in experiments on digitalis (107), cinchona alkaloids (154), analgesics (59, 116), curare (189), antihistamines (10, 11), diuretics (110), etc.

Animal units

At one time it was generally assumed that, if an experiment was carried out under clearly defined conditions, it would always give the same result, so that if the threshold dose of a preparation of a drug was determined then its activity was known. This assumption has led to the definition of a number of so called units of activity in terms of the dose just necessary to produce a standard effect on a standard animal under standard conditions (58, 113). Such units may serve a useful purpose in the early stages of an investigation but it is now recognized that their error may be threefold or more. In 1927 Trevan (186) drew attention to the fact that the threshold dose varies enormously even when the animals are as uniform as possible, and proposed that toxicity tests should be based on the median lethal dose, which kills 50 per cent of the animals. He called this the LD50 (not LD_{50}); the more general term ED50 (median effective dose) is sometimes convenient (96). This proposal introduced a quantity which could be accurately measured and was a great stimulus to further work. The ED50 plays a part not only in accurate assays with standards, but also in experiments without standards. Lethal doses and animal units are now commonly defined in terms of the dose necessary to have a standard effect on 50 per cent of the animals. Various methods of determining this have been described and will be discussed on p. 117. The term LD100 is sometimes used to mean the smallest dose on which no animals were observed to survive, but it should be avoided because it has no precise meaning.

Standard preparations

Even the most careful measurements of effects upon a given colony of animals cannot eliminate differences between one colony and another or between the

effects on the same colony at different times. Trevan gave examples of large variations in the LD50, and so emphasized a fact that was known to some workers at that time and to almost everyone now. These errors can only be controlled by the use of a standard preparation consisting of the drug itself in stable form. The first reputable standard was a preparation of diphtheria antitoxin used by Ehrlich in 1897. An international standard diphtheria antitoxin was established in 1922 by the League of Nations Health Organization, and now international standards for more than forty drugs are controlled by a committee of the World Health Organization known as the Expert Committee on Biological Standardization (58, 104, 113, 151, 185).

A unit is the amount of specific activity contained in a given weight of the standard preparation. The meaning of this definition depends on the meaning of the word "specific", which depends on the circumstances (126, 148).

All the best bioassays now depend on a comparison between the effects of the unknown or test preparation (U) and those of a standard preparation (S). The result is calculated on the assumption that the effects produced by equivalent doses of these two preparations are exactly equal, and the error is mainly due to the fact that, owing to the variability of biological material, this assumption is only approximately true. It is always desirable and often possible to calculate this variability and the error of the result from the internal evidence of each assay. This can only be done when the two preparations are used simultaneously and when care is taken in the design of the experiment.

III. TYPES OF ASSAY

Bioassays are of three kinds:

1. Direct assays.
2. Assays depending on measured effects.
3. Assays depending on quantal effects.

1. *Direct assays*

In these the individual effective dose (I.E.D.) is measured in each animal. For example, digitalis may be injected intravenously until the heart stops and the result depends upon the amount of digitalis necessary to produce this effect. A satisfactory assay depends on a comparison between the average results on two groups of animals, with the standard preparation and the unknown preparation. The results can best be calculated by converting each I.E.D. to a logarithm. The mean of the logarithms of each group of results gives an estimate of the average I.E.D. The difference between two such means gives an estimate of the ratio of the potencies and the standard deviation of the logarithms gives an estimate of the error.

In such tests some of the glycosides in the digitalis do not have time to exert their full action, and their effect is increased if they are given slowly. If a constant volume is injected per minute, strong solutions thus appear less effective than they should. This introduces an error which can be diminished by requiring that the survival time shall lie within defined limits. An alternative method de-

pends on the observed fact that, if a constant volume is injected per minute, log survival time is linearly related to log concentration (37). This enables the calculations to be based on a measured effect.

A similar technique is used to assay curare by the rabbit-head-drop method (190, 62, 46). The drug is given intravenously through a needle until the rabbit can no longer hold up its head; when this occurs the needle is pulled out of the vein and the rabbit recovers. The accuracy is increased by crossing over the rabbits so that on a second occasion those which received the standard receive the unknown and vice versa. The difference between the logarithms of the effective doses is calculated for each animal and the result depends on the mean of these differences. Such experiments provide an admirable opportunity of studying the factors contributing to animal variation. The rate of injection of the curare can be varied over a sufficient range without altering the I.E.D., but if it is much increased the I.E.D. increases like that of digitalis, and, if the rate is very slow the I.E.D. also increases owing to excretion and inactivation of the drug. It is probable that most of the effects of drugs become less when the drug is given too quickly or too slowly, though in some cases excessive speed may introduce other effects as toxic complications.

Similar experiments have been done on man with, for example, injectable anaesthetics (161).

The same principle is involved whenever a drug is given in a fixed daily dose until a definite effect is produced. Reid (67) injected anterior pituitary extracts once a day into cats until glycosuria appeared and calculated the results from a standard curve connecting day and dose. The experiment was arranged as a cross-over test and the good agreement between replicate results shows that it is an accurate method. Hanzlik (112) gave sodium salicylate to 300 men until toxic symptoms occurred and found a wide variation in the toxic dose. Digitalis has often been tested on man in the same way. Comparisons can be made by giving one drug until a steady state is produced and then changing to another drug. By trial and error it is possible to adjust the doses until the change of drug produces no change in the patient; the doses are now assumed equiactive. Such assays have not received much attention from mathematicians and will not be considered in detail here.

2. *Measured effects (graded effects)*

In these assays the effect of the drug on each animal is measured. A familiar example is the assay of a vitamin which increases the weight of rats; the amount of vitamin can be estimated from the increase of weight. Various hormones are assayed by their effect on the weight of individual organs, such as the thyroid or the internal genitals. Vitamin D can be assayed by its effect on the ash content of the bones. Aneurin can be assayed by its effect on the heart rate. Adrenal cortical extracts and chemotherapeutic agents can be assayed by their effect on the survival time. Numerous other examples could be cited.

Sometimes the estimate depends on a subjective comparison of the observed result with an arbitrary scale prepared in advance. For example the amount of rickets in a rat can be estimated by comparing the appearance of the bones under

X-rays with a series of X-ray pictures. Similarly the effect of progesterone on the rabbit's uterus can be assessed by making histological preparations and comparing their appearance with an arbitrary series of histological preparations. In this case the log-dose-effect curve can always be made straight by making equal intervals of effect correspond to equal intervals of log dose.

3. *Quantal effects*

In these assays the effect on each individual animal is not measured and the result depends on the percentage of animals which show some definite positive reaction such as death or oestrous or hypoglycaemic symptoms. The effect on each animal is said to be quantal (92) ("all-or-none"). Such experiments can be interpreted by using the percentage as the measurement of the effect, but the error cannot be calculated from the variations occurring among the animals in a group, but must be calculated theoretically. This means that the error is known more precisely, but the calculations may become much more complicated and a special section will be devoted to them later in this review (p. 114).

Measured effects can always be made quantal by deciding to call the effect positive when it exceeds some arbitrary threshold. In this way it is possible to avoid consideration of the shape of the dose-effect curve, at the expense of some loss of information. Theoretically it may be expected that about half the information will be lost, so that twice as many observations will be needed for any given degree of accuracy (92, 162).

Quantal effects can sometimes be replaced by measurements of the latent period, or the duration, of the effect. For example, the survival time of an animal is generally shorter when the dose is larger. The survival time can thus be used as a measure of the effect of the drug and the accuracy of the result may be increased by this device. This question is discussed on p. 109.

In some cases the effects may be partly measured and partly quantal. For example, some of the animals may survive indefinitely, or some of the effects may be too small to be measured. Methods of dealing with such "truncated distributions" have been discussed by various writers (178, 138, 120, 121) (cf. also p. 110).

IV. THE RELATION BETWEEN DOSE AND EFFECT

Most calculations about bioassays depend on the shape of the dose-effect curve. The dose is more or less under the control of the experimenter and should therefore always be plotted on the horizontal scale, while the effect, being a dependent variable, is plotted on a vertical scale.

In most cases it is best to plot the logarithm of the dose rather than the dose itself, with a scale of actual doses spaced at logarithmic intervals for the convenience of readers. A simple arithmetic scale of doses is more convenient in some cases and its use will be considered on p. 124.

The advantages of using a logarithmic scale of doses are as follows:

1. The results can be plotted when the doses vary over a 1000-fold range or more just as easily as when they vary over small ranges.

2. If the doses are measured as volumes of two different concentrations of the

same drug, the two curves are similar in shape and slope and differ only in position. The horizontal distance between the two curves is constant and equal to the logarithm of the estimated ratio of the concentrations. It is customary among pharmacologists, though not among mathematicians, to call such curves parallel, and this usage will be followed in this review. If the same results are plotted against an arithmetic scale of doses, the two curves often look quite different from one another. The ratio of their slopes at any given height above the base line is then equal to the estimated ratio of the concentrations.

3. The distributions of the individual effective doses and the results of the tests are in most cases "lognormal" (98) (not "log-normal"). This means that their logarithms are normally distributed and should therefore be used in the calculations where possible. When the error is large the use of logarithms is inevitable. It is, for example, obvious that errors greater than +100 per cent are more likely than errors less than −100 per cent, but it is reasonable to expect that errors of +log 2 are just as likely as errors of −log 2. When the error is small the use of logarithms gives the same result as their avoidance, but is desirable for the sake of uniformity. Lognormal distributions are not peculiar to pharmacology, but are commoner than normal distributions in all biological measurements. They have been found, for example, in measurements of blood sugar, blood pressure, weight, pulse rate, reaction times and the number of words in a sentence by G. Bernard Shaw (98).

4. The error of the result of a test is generally proportional to the result itself. If arithmetic scales are used, the standard error can be calculated as a percentage and called the "coefficient of variation". If logarithms are used, this complication does not arise since the error, calculated as a logarithm, is constant. Because of these facts it is convenient to use logarithms in comparing one test with another. If, for example, a biological test is compared with a physical test in which the results are given in quite different units, the difference between the logarithms of the two results on one preparation of drug gives an estimate of the conversion factor and should be constant. If logarithms are used, it is easy to calculate whether this factor does vary more than can be accounted for by the errors of the two tests. Without logarithms the calculations would be difficult if not impossible (91).

5. Dose-effect curves are generally convex upwards. When doses are plotted on a logarithmic scale this convexity is diminished so that the curve becomes straighter (99). In many cases it is justifiable to assume that a straight line does fit the results with sufficient accuracy over the range of doses used. When pushed to the limits, this assumption implies that if the dose is increased indefinitely the effect also increases indefinitely, so that the only limit to the effect which may be produced is the physical difficulty of administering very large doses. It is, however, usually found in practice that, when very large doses are given the effect does not continue to increase as rapidly as the straight line predicts. At the other end of the curve the straight line predicts that a threshold dose will have no effect at all and that smaller doses will have negative effects, increasing indefinitely as the dose is diminished, and this is not what generally occurs.

Although the formula for a straight line makes improbable predictions about the effects of extremely large or small doses, it does serve a useful purpose when its application is confined to a comparatively narrow range of effects. It is such a convenient formula that it is often assumed to express the facts accurately enough.

It is not always possible to determine the shape of the dose effect curve accurately in one experiment and conclusions about it must often be based on the results of a number of experiments. Curves can be constructed by calculating the mean of the effects produced by a given dose on different days. Such curves are likely to be flatter than the real curve (162). It is better to calculate the average value of the shape of the log–dose-effect line. In cases of doubt the best method of determining whether this line is straight or not is to calculate an index of curvature in each experiment and then consider whether all these indices taken together provide significant evidence of curvature. If they do, it is usually possible to straighten the line by using some mathematical function of the effect as first measured, instead of the effect itself. The word "metameter", which is etymologically related to the word parameter (6), has been coined to denote "the measurement, or transformation of the measurement, used in evaluating biological tests" (8). It is best to distinguish between dose-metameters and effect-metameters. The usual dose-metameter is the logarithm of the dose; a familiar effect-metameter is the probit, which is commonly used instead of a percentage because of its convenience. Various other effect-metameters are discussed below (p. 96).

The ideal effect-metameter should fulfil the following conditions:

(1) It should be linearly related either to dose or preferably to log dose.

(2) Its variance should be stable; that is, it should either be constant or vary in a predictable way.

(3) It should be normally distributed or nearly so.

(4) Its variance should be small and the slope of the curve relating it to log dose should be steep.

These conditions will now be considered in more detail. In this discussion Y denotes the effect as first measured and y the metameter.

1. *Linearity*

The conversion of effects, as first measured, into metameters which will be linearly related to log dose depends on the shape of the log–dose-effect curve. When this is known, it can be used to convert measurements of effects into log–dose-equivalents, and these log–dose-equivalents can be used as effect-metameters. The use of this principle is best known in connection with probits, but it applies also to measured effects. The most general method of accomplishing the desired result is graphical. The shape of the log–dose-effect curve is determined in preliminary experiments and plotted. A convenient scale of metameters is laid out at equal intervals along the log dose scale of this curve, which is then used to convert the effects observed in subsequent experiments into metameters. These metameters are likely to be linearly related to log dose. This simple method

is not much more complicated than the direct use of the effects as first measured. Small amounts of curvature, which would scarcely justify elaborate mathematics, can be allowed for by this graphical method.

Most workers prefer a formula to a graph because it is easier to use and because of the air of accuracy which it confers upon their calculations. This preference is less clearly justified in bioassays than it is in physics, since the theoretical curve predicted by the formula seldom fits the data so well that a better curve could not be drawn by eye. Some of the formulae of physics not only fit the data much more accurately than this, but also receive a rational explanation in simple terms. This is unlikely to occur often in biology because there is a large, but not an infinite, number of variables, and the use of a simple formula is justified only by the empirical finding that it fits the facts well enough for practical purposes. In any given case there are generally quite a number of formulae which fulfil these conditions (53).

Measured effects. When the effect is quantal the choice of metameter is simple because the shape of the curves is fairly constant, but with measured effects the curve may have any shape.

Consider first curves which are convex upwards. Morgan (155) encountered such curves in work on the effect of vitamin A on the growth of rats and used the metameter Y^2 to straighten them. This device, which assumes the original curve to be a parabola, was much used in assays of vitamin A, but later abandoned (111). This metameter may be considered a special case of the metameter Y^i and it is possible that other values of i would be more appropriate in other cases. If the curve remains convex upwards when $i = 2$, a larger value of i is indicated, and vice versa.

When the log–dose-effect curve is concave upwards it can be made straighter by the same transformation with $i < 1$. For example, Eisenhart (63) used the metameter $y = Y^{\frac{1}{2}}$ in interpreting experiments on the effect of oestradiol on uterus weight and Bruce, Parkes and Perry used it in assays of ACTH by its effect on thymus weight (44). Other values of i between 0 and 1 may be suitable in other cases. According to Finney (81) quite large variations of i have little effect on the result of the assay, but much effect on the distribution of its error. It should therefore be possible in some cases to adjust i so as to give both linearity and constant variance.

Log Y is a popular metameter which is used when the log–dose-effect curve is concave upwards. This metameter has been applied to measurements of comb length after androgens (64), vaginal smears after vitamin A (164), bone-ash after vitamin D (90), diuresis (142), pressure thresholds and reaction time for pain (109, 121), and various measurements depending upon time (see p. 109).

When the effect is calculated in terms of a difference between treated animals and control animals, zero dose generally has zero effect and the log–dose-effect line becomes asymptotic to the base line. In these circumstances it is often convenient to straighten the line by plotting log Y against log X. This device has been applied to measurements of the latent period for ergometrine (192), analgesic scores (46) and counts of Leishmania in the spleen (46).

It should be noted that, whenever log X and log Y give straight lines of slope b, X and $Y^{1/b}$ will also give straight lines. Assays of the type considered here can therefore also be interpreted with an arithmetic scale of doses and the result calculated in terms of the slope ratio as described on p. 124. The choice between these two methods depends partly on their effect on the variance of y, which should if possible be constant for all values of y.

Log–dose-effect curves often show two curvatures like an S. In most experiments where a drug produces a reversible effect on an isolated tissue, the log–dose-effect curve is S-shaped and symmetrical. A. J. Clark (52, 53) studied many such curves and interpreted them in terms of a reversible combination between drug and tissue governed by the mass laws. On this theory these curves are identical with the titration curve of a buffer, the oxyhaemoglobin dissociation curve, the enzyme-substrate dissociation curves and the Langmuir adsorption curve. Such curves are sometimes called logistic and can be turned into straight lines by means of the logit transformation, which is discussed below in connection with quantal responses. For all practical purposes logits are the same as probits and it is seldom if ever possible to tell which gives the best fit to the observations. The probit is theoretically best in the interpretation of quantal data and the logit is theoretically best if the shape of the curve depends on an equilibrium governed by the mass laws. It is not, however, certain that Clark's theory of the shape of these curves is correct. It is possible that the tissues contain a population of drug-receptors responding quantally like the animals in a toxicity test (89, 93, 197); in this case the probit would be the appropriate metameter. It would be unfortunate if different metameters were used for two groups of curve which are indistinguishable in shape. Probits are well established and it is recommended that they should be used wherever they fit the data. However, the use of either of these transformations involves a knowledge of the maximum response which generally cannot be determined with accuracy, and in practice data obtained from isolated tissues are usually interpreted without any transformation at all. This is justified when only a small part of the dose effect curve is used.

Quantal effects. It was J. W. Trevan (186, 187, 188, 108) who showed that assays based on quantal effects could be made accurate by proper attention to the design and interpretation of the experiment. He used the approximation that the S-shaped curve, obtained by plotting the percentage of positive effects against the dose (or log dose) was practically straight in its middle range. Durham, Gaddum and Marchal (61) applied similar arguments to toxicity tests without a standard preparation and proposed a simple sequential type of test using small groups of animals at first and more animals if the first results were inconclusive. They published tables, based on the binomial distribution, showing the results to be expected when groups of 5, 10, 15, 20 or 30 animals are selected from populations in which the true mortality takes various values.

The mathematical treatment of quantal effects is facilitated by converting percentages into probits by means of suitable tables (68, 83, 88). The probit (26) is equal to 5 plus the normal equivalent deviation (N.E.D.) (92) and is calculated from the theoretical shape of a normal curve whose standard deviation is

one. The N.E.D. is the deviation (from the mean) equivalent to the given percentage of the area of the curve. Gaddum (92) found that when the N.E.D. was plotted against log dose the results could usually be fitted by straight lines, which shows that the logarithm of the individual effective dose (I.E.D.) is normally distributed. This makes it possible to define the whole curve in terms of two parameters—the LD50 and the standard deviation of the logarithm of the I.E.D. (λ). The probit (or N.E.D.) thus provides a convenient metameter for quantal effects and has been much used in recent years, but its history goes back to the work of Fechner in 1860 (cf. 83).

Various other transformations have also been proposed. Finney discusses half a dozen or more (78). The logistic transformation has been used by various authors. The logit, which is based upon it, was defined by Berkson as $\ln (p/(1 - p))$, where ln denotes the natural logarithm (20). The logit for 50 per cent is zero; higher percentages have positive logits and lower values have negative logits. Between percentages of 25 and 75 one N.E.D. is equal to about 1.6 logits, so that the number of logits in an N.E.D. is about equal to the number of kilometers in a mile. Outside this range the ratio increases, but there is no real practical difference between logits and probits in the ranges generally considered (198). Large differences would arise if results were extrapolated, in order, for example, to estimate the dose which would have a toxic effect on 1 in 1 million persons. There is no direct evidence that either metameter would be valid in this case, but most workers would trust probits more than logits.

The use of logits with quantal data implies the assumption that the logarithms of the individual effective doses are distributed in a complex curve slightly different from the normal curve. Using various sets of actual observations Berkson (20) has fitted curves giving minimum χ^2 using this transformation and curves giving maximum likelihood using probits. The former procedure gave the best fit, as judged by χ^2, but, as Armitage and Allen (4) point out, the same result would undoubtedly have been obtained if probits had been used in both cases. There is no evidence which of these two transformations gives the best fit, but either of them is good enough in practice for most purposes. So long as the normal curve is an integral part of statistical theory, the probits which depend upon it are to be preferred in the interpretation of quantal data to other transformations, like the logit, which have no real practical advantages.

Berkson (23) believes that logits have theoretical advantages and quotes physicochemical experiments in which logits give straight lines when plotted against time. It is difficult to see any connection between such examples and dose-mortality curves; the abscissae are an arithmetic scale of time, instead of a logarithmic scale of dose. He also quotes data on enzymes similar to those considered by Clark and discussed above in connection with various other data depending on the mass laws. Berkson points out, in an attack on probits, that individuals may vary from time to time in their response to drugs. This may be interpreted as a variation of "tolerance", but Berkson dislikes this interpretation and seeks to explain dose-mortality-curves in terms of chance alone. He compares the mortality among animals exposed to a drug to the mortality among

targets exposed to bullets; in both cases the mortality is likely to depend on the dose, even if the targets are uniform. It would, however, be unwise to deduce from this that the targets are in fact uniform. The successful use of cross-over tests shows that animals are not uniform, since it depends on the observed fact that measurements on the same individual at different times may vary much less than observations on different individuals. Berkson implies that if the idea of tolerance is rejected, the laws of chance justify the use of logits, but he gives no reasons for this view, which is difficult to understand. The theory put forward by Yule (208) is at any rate clear cut. He considers the case where the chance of a hit in unit time is constant, and death depends on several hits. The same theory may be applied to the case where a suspension of bacterial spores is injected intravenously. Chances associated with the circulation at the time of the injection will determine the distribution of these spores to different sites in the body. If the arrival of one spore at a favourable site is enough to kill the animal, and q is the chance of survival when one spore is injected, q^x will be the chance when x spores are injected and the dose mortality curve will be exponential. If two or more successful spores are needed, Yule's calculations predict the form of the curve, which is similar to that predicted by the use of probits, but not quite the same. This point was discussed in a less precise form by Gaddum (92).

The angular transformation $y = \sin^{-1} p^{\frac{1}{2}}$ has the practical advantage that the weight factor does not vary for different values of y. It was used in another form by Fisher (85) in 1921, and attention has more recently been drawn to it by Knudsen and Curtis and others (88, 134). Between percentages of 20 and 80 this transformation is practically equivalent to probits, but outside these limits this is not so and it is unlikely that the angular transformation represents the facts. It is therefore recommended that this transformation should only be used in these limits, but it is probably more satisfactory to use probits and assume a constant value for the weight factor (0.5) within these limits (92, 153).

The rankit of Ipsen and Jerne (122) is not a rival metameter for interpreting quantal data; it is a method of finding the appropriate value of the probits to use in testing whether measured data are normally distributed, and in assigning normally distributed scores to ranked data.

Probits have met with opposition because of the complexity of the methods of calculation which have been founded upon them, but this complexity is largely unnecessary.

2. *Constant variance*

The calculations are apt to become complicated unless the variance of the effect metameter is constant, or in other words, unless the results are homoscedastic. This variance may be thought of as consisting of two parts, one of which is independent of the drug and is shown by measurements made when the dose is zero. The drug might be expected to introduce another component, but if the variance really is constant over the whole range of observations, this second component must presumably be negligible.

Metameters which fulfil the second condition (constant variance) may or may

not fulfil the first (linearity). In Baker's calculations (12) on the weights of rat ovaries the metameter log Y fulfilled both conditions. In Box and Cullumbine's calculations (41) on survival time the metameter Y^{-1} fulfilled both conditions and the metameter log Y appeared to fulfil the first but not the second.

In such cases no difficulty arises, but in Koch's observations on ovary weight (38) the metameter Y fulfilled the first condition and the metameter log Y fulfilled the second. This means that the effect was linearly related to log dose, but that the standard deviation of the effect was proportional to the effect itself.

If the relation between the standard deviation (s_Y) and the effect (Y) is found to be $s_Y = \varphi(Y)$, the corresponding metameter (y) whose variance is constant and equal to 1 may be found as follows (74). It is approximately true that $dy/dY = s_y/s_Y$. Put $s_y = 1$ and $s_Y = \varphi(Y)$. Then $y = \int 1/\varphi(Y)\, dY$. For example, if $s_Y = kY$, then $y = (\log kY)/k$, and if $s_Y = \sqrt{Y}$, then $y = 2\sqrt{Y}$. This last result has been found useful in calculations on the effect of aneurin on bradycardia in rats (74). It may also be useful when the effect depends on counting such things as cells under the microscope. The standard deviation of that part of the error which depends on the count should be equal to the square root of the total count; and the metameter $2\sqrt{Y}$ should make s constant and equal to 1 (69).

Winder (97) made an elaborate study of the results of tests on analgesics and compared the properties of 12 different metameters, all based on measurements of heat thresholds before (Z) and after (Y) the drug. Eight of these were rejected because the variance varied and one because the distribution was probably not normal. The three remaining metameters (Y, Z/Y, and log (Y/Z)) all gave practically straight and parallel lines with three drugs and there was little to choose between them.

3. *Normal distribution*

It may be argued that the effect-metameter is likely to be normally distributed if condition 1 (linearity) is fulfilled, since the logarithm of the individual effective doses generally has a normal distribution and those effect-metameters which are linearly related to it should also have normal distributions. Unfortunately this argument only applies to that part of the effect-variance which is due to the drug. The assumption is however often true enough for practical purposes and its accuracy is increased when the effect-metameter is the mean of observations on a number of individual animals. For this reason it is important to convert the measured effect on each animal separately into the appropriate metameter before calculating the mean. If the order in which these two steps are taken is reversed, the calculations are simpler, but the distributions are less nearly normal, and the logic of the later parts of the argument is impaired.

One method of determining whether a set of observations is normally distributed is to calculate the third and fourth moments by summing the third and fourth powers of the deviations. Another method is to fit a normal curve to the results and calculate the goodness of fit in terms of χ^2 (87, 175). A third method

is to arrange the observations in order of size, convert them into probits and plot the probit against the measurement. The simplest way to do this is to divide 100 into n parts, allot one observation to the middle of each part, and then convert the n percentages so obtained into probits. If, for example, $n = 5$, the percentages are 10, 30, 50, 80 and 90 and the corresponding probits 3.72, 4.48, 5, 5.52, and 6.28.

It is, however, theoretically better (though it makes little difference) to use another quantity which is given in Table XX by Fisher and Yates (88), and which Ipsen and Jerne call the rankit. This gives the 5 scores as -1.16, -0.5, 0, 0.5 and 1.16. These figures are similar to normal equivalent deviations; if 5 is added to each of them, they are nearly equal to the figures given above. When n is large the rankit becomes even more nearly equal to the N.E.D. calculated by the method given above.

Rankits or corrected probits may be plotted against the measurement. If the distribution is normal, the plotted points should be randomly distributed about a straight line. If the distribution is not normal, they should diverge from a straight line in a regular way and, what is more important, they should always diverge in the same way when the experiment is repeated.

Rankits can also be used when a set of observations is arranged in order of size with no measurement attached to each individual observation. The results of assays are not obtained in this form.

Tests of the normality of the distribution of the response metameters are not made as often as they should be (41, 109, 206).

4. *Small variance. Index of precision*

The error of a test is small when the standard deviation of the effect metameter (s) is small (for a given value of b). This quantity is an estimate of the mean square deviation of the effect metameter from its theoretical value. If the metameters are divided into dose groups, so that all the members of one group are estimates of the same theoretical value, s may be estimated by summing the squares of the deviations from the mean within these groups and dividing by the total number of degrees of freedom ($S(n - 1)$). When the number of animals

TABLE 1

d_n = Range/(Standard deviation)

n	2	3	4	5	6	7	8	9	10
d_n	1.13	1.69	2.06	2.33	2.53	2.70	2.85	2.97	3.08

in each dose group is less than 10, s can be calculated from the range (the difference between the largest and smallest figure). The ratio (range)/(standard deviation) is denoted by d_n and depends on the number in the group (n); this number should always be stated whenever a range is given. If $n < 10$, the use of this factor involves the loss of less than 20 per cent of the information; when the mean of several such estimates of s is used, the error of the calculation is quite negligible.

Convenient methods are also available for calculating the significance of differences between means from ranges (147). This method was first applied to bioassays by Knudsen and Randall (135).

These methods of calculation are easy to understand, but other methods may be more logical and less laborious. When regression lines have been fitted to the data the theoretical value is that obtained from these lines. When precautions have been taken to eliminate the effects of concomitant factors such as litter-origin s must be calculated from the variance within litters and within dose groups. A formal analysis of variance provides the best method of doing this. The application of variance analysis to bioassays was first discussed by Bliss and Marks (39, 40) and subsequent writers have generally adopted their notation and methods (cf. 36, 68).

The total sum of squares for all the observations may be analysed into various components due to the use of different doses, and to concomitant factors such as litter origin. These components of the variance are discussed later. When all such components have been eliminated the residue is called the error, and is used to calculate s. When any given component does not differ significantly from the error component it is included with the error to give an estimate of s based on the maximum possible number of degrees of freedom.

The error of an assay also depends on the slope of the log-dose-effect curve (dy/dx or b). Methods of calculating b are discussed later. The effects of these two factors (s and b) on the error can all be calculated from the ratio s/b. This quantity is an estimate of the standard deviation of the logarithms of the individual effective doses (S.D. log tolerance), for which the symbol λ (lambda) has been used (91, 92, 38). In quantal assays the symbol b has a slightly different meaning, and λ is estimated from $1/b$. The symbol was chosen because it is the Greek form of the first letter of the word logarithm. It is a convenient index of the precision of an assay which is independent of the units in which doses are measured and of the experimental design (arrangement of doses, number of animals, etc.); if λ is known, anyone can predict the error to be expected with any particular design. There is evidence that in some cases at least λ may remain constant when s and b vary (139).

The minimum error to be expected when standard and unknown have equal average effects is approximately given by the following formulae:

$V(M)$ = minimum variance of the logarithm of the result of the assay = $4\lambda^2/N$ (for measured effects) or $8\lambda^2/N$ (for quantal effects) (92).

Another index of accuracy is the weight per animal, which is equal to $1/NV(M)$. The minimum value of this, when $M = O$, in symmetrical assays with measured effects is about $\frac{1}{4}\lambda^2$. When M = the dose range this falls to about $\frac{1}{10}\lambda^2$. With quantal effects the weights are about half these quantities.

For an error ratio ($P = 0.95$) of 1.25 the minimum number of animals is about $1640\lambda^2$ and for an error ratio ($P = 0.95$) of 2 it is about $170\lambda^2$. With quantal effects about twice as many animals will be needed.

The reciprocal of λ is, however, often more convenient, since its sampling distribution approximates to normality when N is large. Woolf (204) has proposed that $1/\lambda$ be called L and points out that, when the response is measured,

the distribution of L for small values of N can be calculated from the fact that b/s_b is distributed as is t (86). To estimate the error of a method from the results of a series of assays it is therefore best to take the mean value of L. It seems likely that L will eventually replace λ as an index of precision. It varies from about 2 for an inaccurate assay to 30 or more for an accurate one (see Tables 2 and). 5 The value of an assay depends very much on the skill of its inventor in discover-

TABLE 2

Indices of Precision

		L	λ	REF.
25, 45 Various		1.1–25.1	0.04 –0.91	(92)
		2–29	0.034–0.5	(38)
19 hormones		2.2–10	0.1 –0.45	(70)
Bacterial toxin	Death. Mice	0.6–1.4	0.72 –1.7	(127)
Virus	Tumours. Chickens	0.84	1.2	(45)
Encephalomyelitis	Infection. Mice	1.25	0.8	(102)
Vitamin D	Rickets. Rats	1.7–2.8	0.45 –0.6	(131)
Cocaine	Anaesthesia. Rabbit	1.3–2.4	0.41 –0.77	(206)
Gonadotrophin	Hyperaemia. Rats	2.1	0.45	(2)
ACTH	Ascorbic acid. Rats	2–6	0.16 –0.5	
Pneumococcus serum	Survival. Mice	2.3	0.44	(127)
Diuretics	Diuresis. Women	2.7	0.37	(110)
	Men	3.2	0.32	(110)
Analgesics	Guinea pigs	3.7	0.27	(197)
Thyroid	Death. Mice	4.5	0.22	(172)
Insulin	Symptoms. Mice	4.6, 4.9	0.2, 0.22	(207)
Salicylates	Symptoms. Man	5.7	0.18	(112)
Ergometrine	Temperature, Rabbits	5.8	0.17	(16)
	Pupil, Rabbits	6.9	0.15	(16)
Gonadotrophin	Prostate weight. Rats	7.2	0.14	(146)
Digitalis	Death 1 hr. Frogs	6	0.17	(152)
	Death 18 hrs. Frogs	10	0.1	(152)
	Cardiac effects. Chick	8.8	0.11	(139)
Sodium amytal	Anaesthesia. Man	8.7	0.11	(161)
Organic arsenicals	Death. Mice	9.7, 14.5	0.07, 0.1	(162)
Diuretics	Diuresis. Mice	13	0.077	(142)
Digitalis	Death. Cat	14	0.071	(32)

ing methods of increasing L. The errors of new tests are often given in terms of the range of error corresponding to some particular number of animals and using one or other of various methods of calculation. While this method of statement gives some idea of the practical value of a test, it would be much easier to compare one test with another if everyone gave their results in the same terms. It is therefore urged that all statements of the errors of tests should include an estimate of L or λ, in addition to any other form of statement which may be thought desirable.

V. CONCOMITANT FACTORS

Since the error of the result of a test largely depends on the variations between animals it is important to cut this down as much as possible. The known causes

of variation, which are sometimes called concomitant factors, may be quantal like sex or graded like body weight and will be considered under these headings. Care should be taken to avoid unnecessary variations in controllable factors like diet and room temperature, but the sensitivity of an animal may depend on a number of other factors which can only be partially controlled.

Precautions must be taken to ensure that the result is not biased by the uneven distribution of such factors among the different groups of animals used in an assay. This may be done either by making sure that the allocation of the animals to the different groups depends entirely upon chance or by taking deliberate steps to ensure an even distribution of the factors which are known to be important. The second method reduces the error, but increases the difficulty of procuring suitable animals.

When a large number of mice arrive in one box to be used in an experiment, it is probable that lazy mice will tend to be caught first. It is theoretically possible that laziness may be correlated with sensitivity to drugs, and it would therefore be unwise to allocate all the first batch of captives to the same dose group.

If the mice are distributed by a process like the dealing out of a pack of cards, some groups will still tend to receive lazier mice than others, although the tendency will be small. It is theoretically better to use a process like the tossing of a penny, since the distribution will then certainly be random and the result unbiased. The use of Latin Squares is really best, since this not only gives an unbiased result, but also, if the laziness of the mice really does matter, it diminishes the error by making sure that the lazy mice are evenly distributed. On the other hand, little is gained by this reduction of error unless its extent is estimated, and this involves some increase in the complexity of the calculations. It should perhaps be added that the example chosen to illustrate this discussion is largely academic; there is no evidence that laziness is at all commonly associated with sensitivity to drugs. On the other hand, Emmens has found a correlation between the weight of mice and the order in which they were caught (68).

The experiments of Chance with amphetamine and allied drugs have emphasized the importance of standardizing all variables. He found, for example, that the effect of these drugs on a mouse was increased by the presence of other mice in the same cage (48, 49).

Quantal concomitant factors

Bliss and Cattell (38) quote a number of cases in which sex has been found to affect the sensitivity of animals to drugs. It is inevitable that this should be so in some cases. Assays depending on the uterus or prostate can only be done in one sex, but less obvious examples are known, in which the sexes differ in their responses to drugs such as vitamin A and squill.

The source of the animals is another important factor. Large differences may occur between animals obtained from different colonies; the variation within litters may be less than the variation between litters (111, 7), though the difference is often small (131, 146). It is sometimes assumed that these differences are due to heredity, but environment may also be very important. Even when two

colonies are supposed to be fed on the same diet and kept at the same temperature there are generally other differences, known and unknown, in the conditions in which they are kept. The animals from one litter have enjoyed not only a common genetic origin but also a common upbringing. The use of the word "isogenic" to mean "from the same sex and litter" (6) may thus be criticized as tendentious. There is little or no direct evidence that genetic homogeneity itself increases the uniformity of the response to drugs. It may even decrease it (65).

Whatever precautions are taken, it is impossible to eliminate completely the differences between one individual animal and another and it is generally desirable, when possible, to assess the effect of each dose of each drug on each individual animal. In skin tests it is often possible to apply all the doses to different areas of skin at the same time and the error then depends on differences between the sensitivities of different areas of skin and is likely, though by no means certain, to be small (177).

In other cases, as in the assay of insulin on rabbits or of drugs acting on plain muscle in acute experiments, it is only possible to test one dose at a time and the error then depends on changes due to time. Time, unlike the other factors considered so far, is of course measurable and could be regarded as a measured cause of error, but it is generally more convenient to consider it quantally since its effect is irregular. Time is divided either into days or into shorter periods, and its effects are eliminated in the design of the experiment.

The best way of allowing for the effects of quantal factors depends on the conditions. When the animals are divided into a few large groups like the two sexes, it may be best to use only one group, or to carry out completely independent experiments with each group and average the results (56). If the same device were used with more numerous groups, such as litters, too many degrees of freedom would be wasted. In such cases the result of the assay may be calculated as if nothing was known about concomitant factors, but allowance must be made for them in the calculation of the error. This may be done by analysing the variance and calculating s from the residual error, after the elimination of all components due to doses and quantal concomitant factors such as those considered here. This method of calculation assumes that the concomitant factors influence the mean sensitivity, but not the shape, of the log-dose-effect curve.

Bliss has described a convenient method of computation which gives the same result more simply, being based entirely on differences within groups (33). This method has been much used in acute experiments on plain muscle (157, 100). The question of litter mates with quantal effects has been discussed by Irwin (126) and Finney (83).

Measured concomitant factors

Weight and age are two important measurable factors which depend to some extent on one another, but may vary independently. Immature animals, for example, are often quite insensitive to sex hormones and show many other peculiarities. In the experiments of Green, Young and Godfrey (109) the pressure threshold for pain was directly proportional to the age of rats up to at least 20

weeks. In these circumstances the threshold per week might be used as metameter, or its logarithm. Another type of measurable factor is the initial reading, in assays where readings are taken before and after the administration of a drug. For example, the fall of blood sugar after insulin depends upon the value of the blood sugar before insulin.

The effect of measured factors may be diminished by restricting their range. It may, for example, be laid down that the animals shall all fall within a certain range of weights or ages; but this makes it difficult to get enough animals unless the range is fairly wide. Attempts are sometimes made to increase the accuracy by adjusting the dose given to each animal according to its weight. Various different methods of doing this have been proposed, but none of them is satisfactory in all circumstances. It is generally best either to make the dose directly proportional to the weight of the animal and to state it in mg. per kg., or to give the same dose to all the animals whatever their weight. The second alternative means much less work, and may be just as accurate.

The best method of dealing with these measurable factors is to correct for them in the calculations. This may be illustrated by considering developments of the method for the assay of insulin by its effect on the blood sugar of rabbits (115, 72, 173, 69). It is usual to determine the initial blood sugar before the injection and also the final blood sugar, which may be the average of results of 5 samples taken at hourly intervals after the injection. The fall of blood sugar depends on the initial blood sugar; when it starts high its fall is greater. The fall was therefore calculated as a percentage, but this correction was not enough. Careful consideration of the relation between the initial value and the fall led to the conclusion that it was better to calculate the fall as a percentage and then subtract three tenths of the initial value measured in mg. per 100 cc. This procedure has been criticized by Emmens (69) who found that equally good results were obtained by neglecting the initial value altogether and basing the calculations directly on the final value. According to him the accuracy gained by the use of the correction was just enough to compensate for the inaccuracy introduced by the unnecessary inclusion in the calculations of the error inherent in the estimate of the initial blood sugar. He suggests the use of the final values with a correction depending on the initial value—a procedure recommended by de Jongh and Laqueur (quoted from 72).

The same problem arises in other cases where the result depends on the difference between two readings, as in assays of ACTH by its effect on the ascorbic acid in the adrenals (68), or of corticoids by their effects on eosinophils. The work of Winder (197) on this subject has already been mentioned (p. 100). In some cases at least, there is no significant correlation between initial and final readings (109). It is then best to neglect the initial readings altogether. It is not clear why Winder was reluctant to do this, even when he found that this procedure gave as good results as any other and better than most.

The use of a correction sometimes has the effect that the metameter becomes zero when the dose is zero. The data for small doses can then be interpreted either by plotting log Y against log X (cf. p. 97) or by plotting Y against X.

If only one scale is logarithmic the curve becomes asymptotic and inconvenient at its lower end. If an arithmetic scale of doses is used the result of an assay depends on the slope ratio and can sometimes be calculated by methods discussed below (p. 124). When this is possible the control data play an essential part in the calculations instead of appearing as a troublesome correction of doubtful value.

Corrections of the kind just discussed can either be discovered by inspiration or calculated by covariance analysis. This means, for example, that the relation between the initial blood sugar and the fall of blood sugar is calculated from the internal evidence of the assays and the appropriate correction calculated from the result. Similar corrections may be applied for other factors such as body weight or age. They may either be calculated once and for all for any particular technique, or estimated from the internal evidence of each assay. Methods of calculation are given in standard textbooks (87, 175). Their application to bioassays has often been discussed (39, 71, 38, 6, 79, 68). If covariance is calculated from the original data and metameters are then used, care is needed in the calculations (74).

VI. THE TIME FACTOR

Measurements of time may appear as part of the dose, or as part of the effect, or as part of both. If different groups of animals are exposed to the same concentration of a toxic gas for different numbers of minutes some of them may die hours later, and the mortality depends on the exposure time, which is thus a measure of the dose. On the other hand, as has already been pointed out, effects may be measured in terms either of the latent period or of the duration of the effect; after large doses the latent period is short and the duration is long. With poisonous drugs the survival time is a measure of the latent period and decreases as the dose increases, but with life-preserving drugs the survival time is a measure of the duration of the effect and increases with the dose.

Sometimes time plays a double role, as it does in measurements of the survival time of goldfish when drugs are added to the water in which they live (105), or in measurements of the latent period for the effects of drugs on plain muscle (192). In these cases time is part of both dose and effect, since the exposure time and the latent period are equal. It is important to bear these distinctions in mind although similar formulae have been used in all cases.

It will only be possible to give here a brief summary of work on this subject. A fuller discussion will be found in A. J. Clark's book (53). Time is generally represented by t which has another meaning in statistics. After some hesitation it has been decided to use t in both senses.

Exposure time

The effect of exposure time has been studied in connection with disinfectants, insecticides and war gases. In all these cases any attempt to express the relation between one drug and another in terms of a potency ratio is liable to be complicated by the fact that the result depends on the exposure time.

1. Large doses are associated with short exposure times and if the concentration (C) is plotted directly against the time required to produce a given effect (t), the results may be fitted by a descending curve which is concave upwards and gives little information.

2. If log t is plotted against log C long stretches of straight line are generally obtained, and the same thing is true when t represents latent period or duration. Log t has also been found useful in connection with the stretching of rubber (195) or muscle (137) under constant load, the fall of surface tension in a solution of saponin (90) and the healing of wounds (58). Some of these data cannot be filled with similar accuracy by any other simple formula, and it is difficult to find a more precise explanation of these facts than that given by du Nouy (158) who suggested that biological time was fundamentally a logarithmic phenomenon.

When the data include counts of quantal effects the relation between probit (y) concentration and time may be represented by the formula $y = a + b_1 \log C + b_2 \log t$. The application of this formula to insecticides has been discussed by various writers and methods have been given for calculating a, b_1 and b_2 (28, 31, 83). The quantities b_1 and b_2 indicate the relative importance of concentration and time in determining the result.

3. It is often convenient to estimate the dose to which an animal is exposed in terms of the product of concentration and time (Ct). When an animal is breathing normally the total amount of a toxic gas inhaled will be proportional to this product, which might thus logically be expected to be an ideal measure of the dose, and was used for this purpose by Haber (163) in work on chemical warfare.

When a cloud of toxic gas mixed with air blows past a sampling apparatus which takes in air at v volumes per minute, the total amount of toxic material collected will be $\int Cvdt$. If this is divided by v the result is an estimate of $\int Cdt$, which is sometimes called the "dose (Ct)" and gives a convenient measure of the total dose. The product of the dose (Ct) and the ventilation rate gives an estimate of the amount of toxic material inhaled.

If the dose (Ct) necessary to produce a quantal effect such as death is plotted against t, it is generally found to be nearly constant over a wide range, but to increase for both large and small values of t. When vertebrates are used this may sometimes be explained by the fact that they hold their breath for a short time in very high concentrations, and detoxicate very low concentrations almost as rapidly as they absorb them, but similar relationships have been found in experiments on the uptake of dyes by simple aquatic plants and in this case there must be some other explanation of the increase of Ct at short times (53).

After considering various ways of comparing the effects of different drugs on goldfish, Gersdorff adopted the minimum value of Ct as a measure of toxicity (105). The same method of plotting has been used by others (42).

4. The formulae which have been discussed so far are only valid over limited ranges, and various more general formulae have been used of which the most popular is

$$(C - C_0)^b(t - t_0) = K \ldots\ldots\ldots\ldots \text{(160, 53)} \quad (1)$$

The value of C_0 can be roughly estimated by plotting C against $1/t$ for large

values of t and finding the value corresponding to $1/t = 0$ by linear extrapolation. The value of t_0 can be similarly estimated by plotting t against $1/C$ for large values of C. The computation of the constants and the use of the formula were discussed by Bliss (31) according to whom t_0 is generally zero, and C_0 is often zero too, so that the methods considered above under 2 can be used.

Other formulae have been found to fit the data more closely in special cases:

$$(C - C_0)(1 - e^{-a(t-t_0)}) = k \quad \text{(106)} \quad (2)$$

$$(C - C_0)t = K(1 + t_0/t) \quad \text{(42)} \quad (3)$$

Latent period

The early history of the use of the latent period as an effect metameter is given by Ipsen (119). In 1904 Arrhenius and Madsen (5) constructed a standard curve connecting the dose of diphtheria toxin with the survival time and used it to convert survival times to dose equivalents. The mean of these dose-equivalents was then taken as a measure of the potency of an unknown toxin.

The principle of converting survival times into dose-equivalents was obviously sound, but in more modern times these dose-equivalents have been used as effect-metameters, in experiments where several doses of standard and unknown samples were used simultaneously. The use of this device has already been discussed in the general section on the linearity of dose-effect lines (p. 95). The complications which arise when some animals survive indefinitely, so that the distribution of survival times is truncated, have also been mentioned (cf. 120, 178).

$y = \log t \qquad x = \log X$

The results can sometimes be fitted by a straight line of the form $\log t = a - b \log X$ which is equivalent to $X^b t =$ a constant. These metameters have been applied successfully to survival times after organic arsenicals (194, 162), the latent period of the action of ergometrine on the rabbit's uterus (192) or of digitalis on embryo chick heart (139), and the speed of action of thrombin (129) and the survival time during infusion of digitalis (37).

Log t has been shown to be normally distributed when t is the induction time for a carcinogen (138).

The data of Smith, Emmens and Parkes (172) on the assay of thyroid by the survival time in closed vessels were analysed by plotting t against $\log X$, but the use of $\log t$ and $\log X$ would probably have produced straighter lines, and does produce more normal distributions (13).

$y = f(t) \qquad x = \log X$

Ipsen (119) found that the formula

$$(X/X_0 - 1)(t/t_0 - 1) = K \quad \text{(4)}$$

(which is really the same as formula (1)) could be fitted to various sets of observations connecting the dose of bacterial toxins with the survival time. Having determined the constants in preliminary experiments, he used this formula to

convert survival times to log–dose-equivalents and calculated the potency ratio by a process which involved the assumption that the curve was constant in shape and slope. If, however, these log–dose-equivalents were used as effect metameters and regression lines fitted, it would be possible to apply tests of validity and to allow variation in the slope of the curve.

$y = 1/t \qquad x = X$ or $\log X$

If the reciprocal of the survival time or "rate of dying" is plotted against the dose of a toxic substance the line which fits the results cuts the base line at a point near the origin corresponding to infinite survival and then rises as the dose increases. It is often linear at first, though it generally flattens out eventually. When, as is often the case, Ct is approximately constant the line given by this method of plotting is a straight line through the origin.

The use of the reciprocal of the survival time has various advantages. If the mean survival time of a dose group is calculated in the ordinary way, a few late deaths may have much too great an influence on the result and survivors can only be included by special methods. If the effect is calculated from the mean of the reciprocals of the survival time, late deaths carry an appropriately small weight and each survivor can be included by adding nothing to the total and one to the number by which it is divided. The question of truncation is thus solved without trouble.

In experiments on the latent period after the injection of viruses, Gard (102) got straight lines by plotting $1/t$ against log dose. These lines presumably become asymptotic for small doses, but the resulting small bend at the bottom of the curve could be neglected and the most constant factor appeared to be the dose corresponding to infinite time. Bryan (45) used the same metameters in experiments on the latent period for tumours and found straight lines, constant slope and constant variance over a 10 million-fold range of doses.

Box and Cullumbine (41) used the metameters $1/t$ and X for data connecting the dose of mustard gas or phosgene with the survival time and found that $1/t$ was linearly related to the dose (not log dose) with constant variance and an approximately normal distribution.

Duration of effect

Measurements of the duration of the effects of drugs differ from other measurements of time in the fact that they increase as the dose increases. When wide ranges of time are used it is generally found convenient to plot $\log t$ against $\log X$. This generally gives straight lines of the form $\log t = a + b \log X$ and the value of b (sometimes called n) gives an index of the slope of the curve which is independent of the units used and has been tabulated for local anaesthetics and other drugs (53, 206). When the range of time is not so wide approximately straight lines can be obtained by plotting t against X or $\log X$ and this method has been used in measurements of the duration of cure after aneurin (140), and melanophore expansion (47) and the survival time after extracts of the adrenal cortex (191).

These results can, however, generally be fitted just as well, and sometimes much better, by plotting log t against log X (36).

VII. REGRESSION LINES AND ERRORS

The simplest and most satisfactory method of estimating the error of an assay is to repeat the whole experiment more than once and calculate the variance of the logarithms of the results from the formula $s_M^2 = S(d^2)/S(n-1)$ where d is the deviation of an individual logarithm from the mean of a group which should all be the same. In the common case where a series of duplicate estimates are available, $s_M^2 = S(d^2)/N$ where d is now the difference between duplicates and N is the total number of assay results (counting each duplicate as two).

In most cases, however, it is possible to estimate the error of an assay from internal evidence as part of the calculation of the result of each assay. Most methods of calculation depend on the assumption that the effect is linearly related to log dose, or in some cases to the dose itself. This assumption is often true, and can usually be made to come true by methods discussed above.

If the effect is plotted on graph paper against log dose, and if a straight line is drawn to fit the observations as well as possible, this line may be assumed to represent the relation between dose and effect. The exact position of such lines is however arbitrary, and a certain amount of error may occur in fitting them. With skill and experience it is usually possible to make this error smaller than the inevitable error of sampling, but most workers wish to avoid this source of error altogether by calculating the position of the line which gives the best possible fit to their results.

Measured effects

There is no difficulty in achieving this object when the distribution of y is normal and its variance constant, as is generally assumed to be the case in tests depending on measured effects. Everyone agrees that the best solution is that given by the method of least squares; the use of this method provides a general principle which is accepted almost axiomatically in cases of this kind. It leads to a minimum value of χ^2. The usual formulae for calculating regression lines are based on this method.

The formula for the regression line may be written

$$y - \bar{y} = b(x - \bar{x})$$

The best estimates of $\bar{x}$, $\bar{y}$ and b depend on the weights (W) of the mean estimates of y. This is calculated differently in different types of assay. For measured effects $W = n/s^2$. For quantal effects $W = nw$, where w is the "weight factor", the value of which depends on the probit and may be obtained from tables (see p. 114).

In either case the constants can be calculated from the following formulae:

$$\bar{x} = \frac{S(Wx)}{S(W)} \qquad \bar{y} = \frac{S(Wy)}{S(W)} \qquad V(\bar{y}) = \frac{1}{S(W)}$$

$$b = \frac{S(Wy(x - \bar{x}))}{S(W(x - \bar{x})^2)} = \frac{S(Wxy) - \bar{x}S(Wy)}{S(Wx^2) - \bar{x}S(Wx)} = \frac{[Wxy]}{[Wx^2]} \quad \ldots\ldots\ldots\ldots (5)$$

$$s_b^2 = V(b) = \frac{1}{S(W(x - \bar{x})^2)} = \frac{1}{S(Wx^2) - \bar{x}S(Wx)} = \frac{1}{[Wx^2]}$$

The second expressions for b and s_b^2 are convenient for computing; the third expressions illustrate a briefer notation meaning the same thing.

In practice an assay generally involves the fitting of two parallel regression lines, and the result depends upon the horizontal distance between them (M). The best estimate of M, the log ratio of the potencies (U/S) is given by the equation:

$$M' = \frac{\bar{y}_u - \bar{y}_s}{b} + \bar{x}_s - \bar{x}_u \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

The value of b is calculated by extending the summations over all the data from both lines. $\bar{x}_s$ and $\bar{y}_s$ are calculated from the results with the standard and $\bar{x}_u$ and $\bar{y}_u$ from those with the unknown.

The accurate formula for the fiducial limits of M' is complicated and has been given in various different notations. Some indication of its derivation is given by the following argument. Consider the quantity

$$\bar{y}_s - \bar{y}_u + b(M - \bar{x}_s + \bar{x}_u) \quad \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

When this is zero M = the most likely value (M'). If the theoretical potency ratio (M) and the doses remain constant, this quantity will be distributed normally about zero over a range of

$$\pm t(A + V(b)(M - \bar{x}_s + \bar{x}_u)^2)^{\frac{1}{2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

where

t is Students t

A = variance of $(\bar{y}_s - \bar{y}_u) = V(\bar{y}_s) + V(\bar{y}_u)$

$V(b)$ = variance of b

$$\frac{1}{V(b)} = \frac{1}{V(b_s)} + \frac{1}{V(b_u)}$$

If the above two expressions (7) and (8) are equated and solved for M, the result gives the fiducial limits of the estimate of M' and is as follows:

$$M = M' + \frac{g}{1 - g}(M' - \bar{x}_s + \bar{x}_u) \pm \frac{t}{b(1 - g)} \sqrt{A(1 - g) + V(b)(M' - \bar{x}_s + \bar{x}_u)^2} \quad \ldots\ldots\ldots\ldots (9)$$

where $g = \dfrac{t^2 V(b)}{b^2}$ = the index of significance of b.

When two solutions are compared which are expected to be equal, and equal doses, measured as volumes of these solutions, are used, then, if weights are assumed constant, $\bar{x}_s = \bar{x}_u$ and this formula becomes simpler.

With measured effects, the number of degrees of freedom for t is that contributing to the estimate of s (125). The variance of probits is calculated theoretically and the number of degrees of freedom is taken as infinite (83).

This is Finney's way of writing the accurate formula for the fiducial limits of assays depending on parallel lines. An equivalent formula was given by Bliss in 1935 (27), but was neglected until its importance was emphasized by Fieller (71) and by Irwin (125), who gave a full discussion of it. It is a special case of a more general formula due to Fieller (73).

The formula has also been given in the following form:

$$M = \bar{x}_s - \bar{x}_u + \frac{C(\bar{y}_u - \bar{y}_s)}{b} \pm \frac{t\sqrt{C}}{b}\sqrt{A + \frac{V(b)C(\bar{y}_u - \bar{y}_s)^2}{b^2}}$$

where

$$C = \frac{b^2}{b^2 - t^2V(b)} = \frac{1}{1 - g} \quad \ldots\ldots\ldots (124, 162) \quad (10)$$

An alternative method of calculation depends on the analysis of variance of the effects. In the notation used by Bliss (34, 36) B^2 is the variance due to slope, D^2 is the variance due to preparation, I is the log dose interval, and k is a constant which takes the value 1, $\sqrt{8/3}$ or $\sqrt{5}$ when 2, 3 or 4 doses of each preparation are used. The following equations may be used.

$$M = \bar{x}_s - \bar{x}_u + kID/B \quad \ldots\ldots\ldots\ldots\ldots\ldots (11)$$

$$b/s_b = B/s \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$C^2 \text{ (Bliss)} = C \text{ (other authors)}$$

$$\text{Fiducial range Bliss} = \bar{x}_s - \bar{x}_u + \frac{C^2(\bar{y}_u - \bar{y}_s)}{b} \pm t\lambda C\sqrt{\frac{1}{S(nw)_s} + \frac{1}{S(nw)_u} + \frac{(\bar{y}_u - \bar{y}_s)^2}{B^2 - s^2t^2}} \quad \ldots\ldots (13)$$

For measured effects the weight factor $(w) = 1$, and for quantal effects $s^2 = 1$.

The same formula (9, 10, or 13) may be used to calculate the fiducial range of an estimate of the log dose (x) corresponding to a given effect (y') on a single curve, except that in this case, M, M', $\bar{x}_s$, $\bar{y}_u$ and $1/S(nw)_u$ become x, x', o, y', and o, respectively.

The fiducial limits are not symmetrically situated about the most likely value. When the observed mean effect of the unknown $(\bar{y}_u)$ is greater than that of the standard $(\bar{y}_s)$, the upper fiducial range is larger than the lower range and vice versa.

When $g < 0.1$ it can be neglected and the general formula for the fiducial limits becomes

$$M = M' \pm \frac{t}{b}\sqrt{A + (\bar{y}_s - \bar{y}_u)^2V(b)/b^2}$$

or

$$M = M' \pm \frac{t}{b}\sqrt{A + V(b)M^2}$$

when $\bar{x}_s = \bar{x}_u$.

The equivalent of this formula was first given by Gaddum (92), and has formed the basis of most of the methods which have been used for calculating errors, but it is not a very good formula and may give misleading results when $g > 0.1$. Much attention has been devoted to small errors due to the use of approximations to the maximum likelihood solutions for quantal effects, and the much larger errors due to g have been neglected.

Quantal effects

It is possible to fit regression lines connecting probit and log dose by methods similar to those discussed above. These methods were originally based on the general principle which is embodied in the method of least squares and are used without hesitation in tests depending on graded effects.

When the data are quantal the question is more complicated. In the first place, the variance of probits is not constant, but varies in a predictable way with the probits themselves, and in the second place, it is doubtful whether the method of least squares can be used at all. The theoretical basis of this method depends upon distributions which are either normal or so nearly normal as not to matter. It is generally agreed that this requirement can be liberally interpreted, but the distribution of probits is grossly abnormal. The histogram showing the distribution of observed probits corresponding to any given true probit is an odd shaped discontinuous asymmetrical curve stretching out to infinity without ever reaching the base line; its variance is infinite. There is no justification for applying the method of least squares blindly to distributions of this kind, and it is therefore necessary to use some other general principle for deciding which line is most likely to represent the relation between dose and effect. The only other widely accepted principle is that embodied in the method of maximum likelihood; the method of least squares is the special case of this method where the distributions are normal.

It is agreed by most though not all (20, 24, 83) writers on this subject that the method of maximum likelihood gives the best solution, but opinions are divided on the question of how much time should be spent on the calculations. At least five methods of calculating regression lines and errors are available. These mostly involve the weight factor w, which is calculated from the variance of p multiplied by the appropriate differential. When n is large the variance of a probit is approximately $1/wn$ or $\frac{p(1-p)}{n}\left(\frac{dy}{dp}\right)^2$. Thus $w = \left(\frac{dp}{dy}\right)^2/p(1-p) =$ the weight factor. Values of w corresponding to any given value of p can be obtained from tables (92, 26, 88, 68, 83).

The five methods of fitting regression lines are as follows:

1. The simple graphical method is accurate enough for many purposes (143) and should always be used first even when other methods are used later. The observed percentage is converted into a probit and plotted against log dose, or logarithmic probability paper is used.

The fact that no correction is made for grouping may be justified as follows.

Imagine all the animals in a dose group ranked in order of tolerance, and all the animals in the population from which they were selected ranked in the same way. Consider the case where the observed result is 2/10 (2 dead out of 10). The second animal in the dose group is likely to be a typical member of the second 10 per cent of all the animals. It is reasonable to place it in the middle of this range at 15 per cent, and to place the third animal at 25 per cent. The third animal may be half-dead, and the true mortality is therefore likely to lie between 15 and 25 per cent, and can thus be expected to be at 20 per cent, which is the value given by first thoughts without any correction for grouping.

If the plotted points are randomly distributed about a straight line the assumption that the individual lethal doses are lognormally distributed is confirmed. If the points seem to lie on a simple curve this assumption is suspect, and if similar curves are regularly obtained, suspicion hardens into conviction. If the line is straight enough the LD50 and the standard deviation of the logarithms of the individual lethal doses (λ) may be determined from the graph. The difference between log LD84.1 and log LD15.9 is an estimate of 2λ.

Lode (145) described a small metal machine for plotting the standard errors of probits without any calculations at all.

de Beer (15) has suggested various ways of simplifying the interpretation of graphs, including a transparent plastic circle with which the slope (or λ) can be read directly. He has also provided nomograms and tables for calculating the limits of error, but he neglects both g and $V(b)$, and this can only be safely done when n is large.

Litchfield and Wilcoxon (144) have described a method of fitting regression lines and calculating errors without using either logarithms or probits; slide rules are permitted but not compulsory. The lines are fitted by eye on probability paper; points depending on observed percentages of 0 and 100 are then added; the line is refitted by eye and tested for goodness of fit and parallelism with nomograms. If the results of these tests are satisfactory, the potency ratio and its error (neglecting g) are calculated by simplified means. All the calculations can be done by an expert in 10–15 minutes.

Finney (84) found that when unpractised draughtsmen tried to use this method the error of fitting was almost as large as the inevitable error of sampling. This should not deter others from acquiring the art of graphical fitting, but they should make independent estimates of their own error before trusting their results.

2. Gaddum (92) believed that the method of maximum likelihood provided the best formula of the line, but realized that this could only be obtained by successive approximations (an "iterative" method). He was deterred from describing how to do this by the belief that the extra accuracy obtained would be negligible compared with the inevitable error of sampling and later work has shown that this is so. He made the surprising discovery that the curve connecting likelihood and probit was very nearly identical in shape with the normal curve. If this approximation were exact, the maximum likelihood solution could be exactly calculated by fitting regression lines with weights calculated from the observed probits and Gaddum accordingly recommended this procedure,

which is much simpler than some of the methods described later and has a negligible error. He also provided a method of using the results when the percentage of positive responses is 0 or 100. The corresponding probits are infinite, but appropriate finite values depending only on the number of animals used may be obtained from a graph or table and plotted directly, or used in the computation of regression lines without preliminary fitting. This procedure gives satisfactory results, but has been little used.

3. Bliss (26) recommended a method of fitting regression lines in which the weight of each observation is calculated from the expected probit, calculated from less accurately fitting lines. If this process is repeated, it eventually leads to the solution given by minimizing the weighted sums of the squares of the deviations, calculated in probits. This quantity is sometimes taken as equal to χ^2, but is not exactly the same thing (126). Gaddum (94) pointed out that the results do not lead to maximum likelihood. In an appendix to this paper by Bliss (26), Fisher described the method of dealing with percentages of 0 and 100 which is now generally used.

4. The best expression for χ^2 is

$$S\left(\frac{(p - p')^2}{p(1 - p)}\right)$$

where p is the value corresponding to the regression line and p' the observed value of the proportion of quantal effects (184, 92). The solution obtained by minimizing this quantity can be obtained by an iterative method (21, 4).

5. An iterative method leading to maximum likelihood was devised by Fisher and described by Bliss (29). Both probits and weights are corrected at each stage of the calculations. First Garwood (103) and then Cornfield and Mantel (55) described other methods which lead to the same result. The theoretical basis of these methods has been discussed and appropriate methods of computing have been described by various writers (88, 68, 83, 46).

TABLE 3

	SLOPE (b)	LOG LD50 (m)
2. Simple method	2.222 (−0.024)	$\bar{3}$.1526 (+0.0010)
3. Bliss' first method	2.299 (+0.053)	3.1484 (−0.0032)
4. Maximum likelihood	2.246 ±0.294	3.1516 ±0.049

Some of these methods were compared by Irwin and Cheeseman (127). Table 3 shows the results of some of their calculations. The figures in parentheses give the discrepancies between the first two results and the last result, which are smaller for method 2 than method 3. In both cases, however, they are much smaller than the inevitable standard error of sampling given in the last line. There are many methods of calculation which produce as good results as this (4, 22, 153) and there is no evidence against the view that the calculation of successive approximations is a pure waste of time. Armitage and Allen (4) fitted curves by both maximum likelihood and minimum χ^2 to a dozen sets of data using probits and logits and the angular transformation and also the methods of

Kärber, Reed and Muench, and Thompson. It is a pity they did not use the simple method (2) recommended above. This has been applied to their data and found to agree with the maximum likelihood solution better than any of the other methods do.

It is commonly assumed that no animals will die when no drug is given, but this is seldom quite true, and it may sometimes be necessary to apply a correction for the "natural response rate". The appropriate formula is obtained as follows. Let p be the proportion dead on a given dose and p_0 the proportion dead on no dose. The proportion killed by the drug is $p - p_0$, out of a maximum possible proportion of $1 - p_0$. The corrected value of p is thus $(p - p_0)/(1 - p_0)$. This is known as Abbott's formula (1), but was first used by Tattersfield and Morris (179).

Finney has described methods for calculating the maximum likelihood solution with this correction (76, 80, 83).

VIII. EXPERIMENTAL DESIGN

Much depends on the design of an assay; good designs save time and give accurate and reliable results with the smallest possible number of animals. The requirements of a good design are to some extent incompatible with one another. The experimental technique and the calculations should both be simple; this means that the doses should be few. The assumptions should be few, and, except when it is known that S and U are qualitatively the same, these assumptions should be tested for validity. This is impossible if the doses are too few, but in routine assays it is often unnecessary.

Each test should be self contained and its error should be estimated from internal evidence. In parallel line assays it is generally best to allot doses in a geometric series so that there is a constant interval between their logarithms, but in slope ratio assays the doses are generally allotted in an ordinary arithmetic series. When two or more samples of drug are compared they should generally be given to the same number of animals. When k unknowns are simultaneously compared with one standard it is theoretically best to allot $\sqrt{k}$ times as many animals to the standard as to each of the unknowns, but it is doubtful if it is worth applying this fact in practice (74).

Apart from these generalities the design depends on the assumptions which can safely be made and on the accuracy with which the result can be foretold.

Experiments without a standard

In experiments to determine the LD50 of a new drug or to estimate activity in animal units it is generally necessary to test a wide range of doses and it would be wasteful to allot many animals to each dose group.

A simple method which is economical in animals but not in time is known as the "up and down" or "staircase" method. The drug is tested on one animal at a time and the result of each experiment determines the dose used in the next experiment. This method is only suitable when the effect is quick; it was originally used in experiments to determine the greatest height from which ammunition could be dropped without exploding (153, 83).

It is more usual to test each of a series of doses simultaneously, but in the first experiment each dose may be given to one animal only, and the ED50 estimated from the mean of the logarithms of the smallest effective dose and the largest ineffective dose (92).

Sometimes larger numbers are used on each of a series of doses covering the whole range. In one such experiment, for example, a series of doses of a bacterial toxin was injected into groups of 5 mice (127, 128). The doses used varied from $\frac{1}{16}$ mg. to 4 mg., each dose being twice the previous one. The numbers killed by these doses were 0, 0, 2, 1, 5, 5, 5.

The LD50 and its error may be calculated from such results by various simple methods which do not involve any consideration of dose-effect curves. These methods have been discussed by various writers (92, 18, 127, 128, 193, 51, 4, 55, 83). Miller (153) divides them into two groups. The first group contains the "double integration" methods associated with the names of Dragstedt and Lang (60), Behrens (17), Reed and Muench (165), and Wright (205). Their theoretical background is quite illogical (196). The number of animals dead on each dose is increased by the addition of the number dead on smaller doses, and the number alive is increased by the addition of the number alive on larger doses. These manoeuvres are thought to be justified by the fact that the additional animals would have reacted in these ways if they had all received the same dose. Corrected estimates of the proportion dead (p) are thus calculated and plotted against the dose or log dose; the LD50 is calculated from the resulting curve. The corrected estimates are not, however, estimates of the true value of p. The first manoeuvre is likely to increase p by an amount depending on the total number of animals receiving smaller doses, and the second manoeuvre is likely to decrease p by an amount depending on the number of animals receiving larger doses. At the LD50 these two effects are likely to balance, provided that the doses are equally spaced and the animals equally distributed. For larger doses the effect of the first manoeuvre is likely to predominate, so that p is overestimated, and vice versa. The resulting curve will thus be not only smoother than the observed curve, but also steeper than the true curve. This method has a limited use as a means of estimating the LD50.

The second group of methods consists of the "moving average" methods associated with the names of Spearman (176), Kärber (132) and Thompson (182, 183). These are generally preferred because they are based on sound logic and because their error can be calculated (128, 193).

In all these methods, except that of Thompson, it is important to be sure that a sufficient range of doses has been covered, and that the next larger dose would have had an effect on every animal and the next lower dose on none. It is impossible to know this without using an undue proportion of animals on extreme doses; and therefore this design is not suitable when accurate estimates are required. It is very suitable for preliminary experiments without a standard preparation and once the experiment has been done quite simple formulae give as good an estimate of the ED50 and its error as any which could be obtained from the data. When more accurate results are needed a standard preparation is used and the experimental design considered here is replaced by other designs depending on

the dose-effect curve, and using doses in the middle range where their effects are likely to have more weight.

The best of these methods is that described by Thompson. The curve is smoothed by averaging the effect of each set of 3 successive doses and plotting the result against the middle dose, and the LD50 is determined by interpolation.

(*1 and 1*) *dose assays*

Theoretically the most accurate assay would be one in which one dose of each preparation (*S* and *U*) was used and they had exactly the same effect in the steepest part of the log-dose-effect curve. Such an assay, however, would provide no estimate of error and it is impossible to get such results regularly because of the sampling error and for other obvious reasons. In practice there will generally be a difference between the two effects. At one time allowance was made for such differences by using a standard dose-effect curve, but this practice has been abandoned by most workers because the slope of the curves is liable to vary. This slope should therefore be determined in every assay and this means that at least two doses of one of the preparations must be used. A standard curve was, however, used in one recent paper with apparently satisfactory results (2). The estimate of the error of the results was actually too small to be compatible with the formulae recommended in this review, even if the curve *was* constant in position.

In routine toxicity tests it is sometimes not necessary to estimate the potency in each experiment, but only to exclude occasional toxic samples of a known drug. In such cases it is justifiable to give a dose (x) of the unknown preparation to one group of animals, and a larger dose (kx) of the standard preparation to another group of animals, and to stipulate that the unknown preparation will be rejected if it has more effect than the standard preparation in spite of the fact that it is used in a smaller dose. Methods have been devised for predicting the effects of such tests on preparations with different toxicities relative to the standard (97, 162). If the slope of the log-dose-mortality-curve (L or b) is known, a curve can be drawn showing the relation between the toxicity and the percentage of rejections, but even when L is not known the general effect of such tests can be foreseen. It is obvious that when the toxicity of U is 100 k per cent of that of S, the probability of a pass is 0.5, with a small correction for grouping. It is thus possible, without knowing the slope of the dose-effect curve, to define exactly the level of toxicity necessary to ensure 50 per cent of rejections.

The best results are of course obtained when the animals are as homogeneous as possible. If heterogeneous animals are used, the test may become somewhat more lenient to toxic batches, but only at the expense of rejecting good batches.

Therapeutic activity can be ensured by similar tests with $k < 1$. The biological tests of neoarsphenamine in the British Pharmacopoeia are based on these principles.

(*2 and 1*) *dose assays*

The simplest acceptable design for an assay is one in which one dose of the unknown preparation has an effect intermediate between those of two doses of

the standard preparation. If it can be assumed that the log–dose-effect curve is straight, the calculations are very simple and the result may be obtained graphically, since it is only possible to draw one straight line through two points. This design provides no test of the validity of the assumption on which it is based, but is useful in routine tests because of the simplicity of the calculations. The formulae for calculating the results of (2 and 1) dose assays in Table 4 are exact. The other formulae involve the assumption that the weights of the results from different dose groups are equal.

The error of the potency estimate is least when all three percentages are near 50. This cannot be, unless the log dose interval for the 2 doses is small, and in this case there is a danger (1) that the effect of the one dose will not lie between the effects of the two doses and (2) that the slope will not be significantly greater than zero. Either of these misfortunes will mar the result and it is therefore necessary to steer between the dangers associated with large and small dose intervals.

When the effects are quantal, the slope (b) is estimated with maximum accuracy when the percentages are about 6 and 94 (92, 83).

(2 and 2) dose assays. Four-point assays (8)

This experimental design was first discussed by Gaddum (92) who pointed out that it could be applied either to quantitative or to quantal tests, and found that if the weights of the different estimates of the effect were assumed equal, the regression formulae took a simple form which could also be deduced directly from simple principles. This formula has been much used and nomograms have been constructed to aid the calculations (135). In quantal tests the weights are generally not equal, but the error introduced by this fact is small.

The differences between the mean effects of the 4 doses provide 3 degrees of freedom which may be used to estimate E, F and G (Table 4). E and F give the result of the assay; G provides a test of the validity of the assumptions on which it is based. If G does not differ significantly from zero it is reasonable to be content. The observation that this was so in a series of routine tests of vitamin A confirmed the soundness of the technique (111). A significant value of G can be interpreted as evidence that the log–dose-effect lines are either not parallel or not straight. If they are not parallel, the result is useless. On the other hand, the results of a (2 and 2) dose assay can always be fitted with two parallel parabolas, and it so happens that the formula for calculating the result of the assay on this assumption is identical with that used on the simpler assumption of straight lines. This interesting fact shows that the formula may give the right result even when the slopes differ, but in these circumstances it must be used with caution (75, 111, 199).

More complex designs

When more than (2 and 2) doses are used, it is possible to distinguish curvature from lack of parallelism, and to apply tests for opposed curvature, double curvature, etc. For example, the results of a (3 and 3) dose assay provide two quantities

TABLE 4

Balanced parallel line assays

		(2 AND 1) DOSE	(2 AND 2) DOSE	(3 AND 3) DOSE	TWIN CROSSOVER
Mean effects					
Unknown		U	U_1 , U_2	U_1 , U_2 , U_3	—
Standard		S_1 , S_2	S_1 , S_2	S_1 , S_2 , S_3	—
Group differences		—	—	—	y_1 , y_2 , y_3 , y_4
Dose difference	E	$S_2 - S_1$	$\frac{1}{2}(U_2 - U_1 + S_2 - S_1)$	$\frac{1}{4}(U_3 - U_1 + S_3 - S_1)$	$\frac{1}{4}(y_1 - y_2 - y_3 + y_4)$
Preparation difference	F	$U - \frac{1}{2}(S_1 + S_2)$	$\frac{1}{2}(U_1 + U_2 - S_1 - S_2)$	$\frac{1}{3}(U_1 + U_2 + U_3 - S_1 - S_2 - S_3)$	$\frac{1}{4}(y_1 + y_2 + y_3 + y_4)$
Slope difference	G	—	$U_2 - U_1 - S_2 + S_1$	$\frac{1}{2}(U_3 - U_1 - S_3 + S_1)$	—
Log ratio of doses	I				
Slope	b		E/I		
Log potency ratio (U/S)	M'		F/b		
Calculation of errors					
Variance of each mean	V	s^2/n *or* $1/wn$	s^2/n *or* $1/wn$	s^2/n *or* $1/wn$	$s^2/2n$
Variance of F	A	$3V/2$	V	$2V/3$	$V/2$
Variance of b	$V(b)$	$2V/I^2$	V/I^2	$V/4I^2$	$V/2I^2$
Variance of G		—	$4V$	V	—
Index of significance of b	g		$V(b)\, t^2/b^2$		
Fiducial limits of M			$M' + \frac{gM'}{1-g} \pm \frac{t}{b(1-g)} \sqrt{A(1-g) + V(b)M'^2}$		
Limits of M when $g = 0$			$M' \pm \frac{t}{b} \sqrt{A + V(b)M'^2}$		

The potency ratio is the ratio of the estimated potency of U to the assumed potency. In most cases the assumed potency is the result when corresponding doses of S and U have equal effects. In (2 and 1) dose assays it is the result when the log dose of U is equivalent to the mean of the two log doses of S.

representing joint curvature ($U_1 + U_3 - 2U_2 + S_1 + S_3 - 2S_2$) and opposed curvature ($U_1 + U_3 - 2U_2 - S_1 - S_3 + 2S_2$). Each of these can be compared with its standard deviation ($\sqrt{12V}$), and if it is significant the calculations are not valid. Such tests for invalidity may be used in initial experiments to discover the shape of the curve. The curve can then generally be straightened by the use of a suitable metameter, and once this has been found it will generally be effective in routine tests. It is however best to make sure that this is so, and that the results really do lie on the straight part of the curve. This is more likely to be achieved by keeping the effects in the same range than by keeping the doses in the same range. Some workers use (3 and 3) doses in routine tests and apply tests for curvature in every experiment. Convenient formulae and a nomogram have been provided (171, 114).

The full calculations with an indefinite number of doses have been described by various writers (71, 68, 46, 36).

IX. ACUTE EFFECTS ON PLAIN MUSCLE

In experiments on plain muscle or on the blood pressure it is often possible to obtain effects at intervals of a minute or two, and two preparations of drug may be compared by adjusting the doses until they have equal effects. In such tests the error is due to variations in the sensitivity of the same piece of tissue at different times and is diminished if the result is based only on comparison between effects produced at nearly the same time.

Schild (69) was the first to discuss in detail the application of mathematics to this kind of test. He proposed the use of a (2 and 2) dose design in which large and small doses of standard and unknown samples are given in random order and the effects measured. The same group of 4 doses is given repeatedly using a fresh random order each time. The variance of the effects is analysed into components due to time (*i.e.*, group of effects), sample, slope, slope difference, and error. The variance of the result of the assay is calculated from the error term, which is equivalent to assuming that the effect of time can be eliminated by adding the same correction to all the effects in one group whether they are large or small. Interactions are neglected so that, for example, the slope is assumed constant. This technique has been used without much change by various other workers (118, 57, 100), and adopted in the British Pharmacopoeia.

The order of the doses is sometimes arranged by means of a Latin Square (174, 136). This means that the duration of the assay must be decided before it is known how long the preparation will give steady results. With this design it is possible to eliminate variance due to columns (*i.e.*, due to the position of each dose in the series of 4 doses in each group), but it would probably be more appropriate to eliminate a component depending on whether each dose was preceded by a small dose or a large dose.

Table 5 shows some values of L and λ for assays of this kind. The values of L are generally high, as was to be expected from the fact that they are calculated from effects on the same piece of tissue kept under constant conditions.

An alternative design for such experiments is the constant-standard method.

It might be said to have been more or less in use for many years before 1943 when Vos (192) gave it a precise mathematical form. The two preparations are given alternately at a constant time interval. The dose of the standard is kept constant and that of the unknown is varied so that some of its effects are larger and some smaller than those of the standard. Effect-differences are calculated by subtracting from the effects of the unknown the mean of the two nearest effects of the standard. The regression line connecting these effect-differences with the difference between log doses of standard and unknown is calculated. The result corresponds to the point on this line where the effect-difference is zero. This method was applied by Thompson to assays of posterior pituitary extracts (180) and

TABLE 5

Indices of precision

Comparisons on the same animal in acute experiments

DRUG	ANIMAL	TISSUE	L	λ	
(a) (2 and 2) dose assays					
Adrenaline	Dog	Blood pressure	5	0.2	(117)
Posterior pituitary	Guinea-pig	Uterus	5.1	0.2	(156)
Ergometrine	Rabbit	Pupil	7	0.14	(16)
Adrenaline	Rat	Colon	8	0.13	(100)
Adrenaline	Rat	Uterus	16	0.062	(100)
Posterior pituitary	Rat	Uterus	18	0.057	(118)
Posterior pituitary	Chicken	Blood pressure	23	0.043	(174)
Adrenaline	Dog	Blood pressure	26	0.038	(136)
Adrenaline	Dog	Blood pressure	30	0.033	(157)
Histamine	Guinea-pig	Ileum	30	0.033	(169)
(b) Constant standard assays					
Ergometrine	Rabbit	Uterus	23	0.044	(192)
Posterior pituitary	Chicken	Blood pressure	32	0.036	(180)
Adrenaline	Cat	Blood pressure	34	0.029	(181)
Adrenaline	Dog	Blood pressure	43	0.023	(181)

epinephrine (181) and was adopted in the U. S. Pharmacopoeia XIV (1950). It has been criticized as laborious, but is likely to be more accurate than (2 and 2) assays since the sensitivity is less likely to be disturbed by changes of dose, and since the effect of the unknown is compared with the effect expected from the standard at the same actual time and not only at about the same time. The figures given in Table 5 support the view that this is a particularly accurate method. The error can be calculated, but since only one dose of standard is used it is not possible to test for difference of slope.

X. CROSS-OVER TESTS

The difference between one animal and another can sometimes be eliminated by testing both drugs on the same animal. Except when both tests can be made

simultaneously this means the introduction of a new variable—time. In some cases variance due to time is at least as large as variance due to differences between individual animals (109, 206). Nothing is then to be gained by using each animal more than once, unless the experiment is designed as a cross-over test. This device was used by Marks (150) in 1925 for the assay of insulin by tests on two days. On the first day some animals receive the standard preparation and some the unknown, and on the second day the preparations are crossed over so that eventually each animal receives both preparations. The result is calculated from the difference between the two results with each animal, which eliminates variance due to animals. Variance due to time disappears in the calculation of the mean results with each dose group, since half of these results were obtained on each day.

The analysis of the variance of the results of such tests was discussed by Fieller (71). If there are $4n$ measurements of the effects on $2n$ animals, the $4n - 1$ degrees of freedom may be allotted to doses (1), days (1), animals ($2n - 1$), and error ($2n - 2$).

The original cross-over test provided no evidence about the shape of the dose-effect curve, and the results were interpreted with a standard curve. Various modifications of the test have been devised to overcome this defect, and the most satisfactory of these is known as a twin cross-over test (173). Each part of this test may be regarded as a (2 and 2) dose assay, but those animals which receive one preparation in one part of the test receive the other preparation in the other part, and those which receive low doses in one part receive high doses in the other part of the test. The result is calculated from the differences between the two effects on each animal.

The average differences in four groups of animals are denoted by the symbols y_1 y_2 y_3 and y_4 and these are respectively equal to $(U_2 - S_1)$, $(U_1 - S_2)$, $(U_1 - S_2)$ and $(U_2 - S_1)$. The meaning and use of these symbols are shown in Table 4. s^2 is calculated from the variance of the observed values of y_1, y_2, y_3 and y_4 within the four groups of animals, and is equal to $S(d^2)/(N - 4)$, where d is the difference between any given value of y and the mean of the group to which it belongs and N is the total number of animals. The analysis of variance in such tests was discussed by Finney (46). The only available test of validity is a particularly inexacting one.

The same principle may be applied to direct assays in which the effective dose is measured in each animal, as in the assay of curare by the head-drop method with rabbits (p. 92). In this case the calculations are comparatively simple.

XI. ARITHMETIC SCALE OF DOSES

Slope ratio assays

It is sometimes assumed that the effect of a drug will always be proportional to the dose, but this is not always so. The best known examples of this type of relation are found among microbiological assays, in which substances are estimated by their stimulant effect on the growth of microbes in an incomplete medium. Some such experiments are best interpreted by the methods discussed

above (9), but sometimes it is found that if the response of the microbes is plotted directly against the dose of the substance, the results lie approximately on straight lines. If the doses are measured as volumes of two different dilutions of the same substance, the two straight lines often both pass through the same point corresponding to zero dose, and the estimate of the ratio of the dilutions is equal to the ratio of the slopes of the two lines. Methods of designing such assays and calculating their results and errors were first discussed by Wood and Finney (199, 75, 200, 77, 203, 202, 83).

The simplest kind of experiment depends on measuring the effects of one dose of each preparation and the effect of no dose at all. This is known as the common zero 3-point design. It gives the most accurate possible result if all the assumptions are correct, but provides no means of testing whether this is so or not and is therefore rather risky. It should only be used in routine tests when much experience has shown it to be justifiable. A 4-point design with 2 doses of each preparation may be used when the dose-effect lines are not quite straight at their lower ends, but this design is not recommended. A 5-point design with 2 doses of each preparation and one zero dose is the best design for general use.

General discussions of the interpretation of such experiments have been published by various writers (203, 35, 68, 46, 36, 82, 54). The most general method depends on the analysis of variance, but the formulae in Table 6 give the same results. These formulae have been simplified, as in the case of parallel line assays (Table 4), by assuming that the weights of the results with different dose groups are equal. On the other hand the formulae have become rather complex because they give the exact fiducial limits worked out from Fieller's formula (73). If $g < 0.1$ it can be neglected and the formulae become much simpler.

The dose scale is given in arbitrary units. Absolute values of b_u and b_s can be obtained by dividing the values from the table by the maximum dose, but these values are not generally required.

Table 6 gives various methods of testing for validity, all of which depend on V and involve the consultation of a table of t, using the number of degrees of freedom which was used to estimate V, *i.e.*, $S(n - 1)$. The index of significance of the slope (g) provides one kind of test; if $g > 1$, b_s is not significant at all and the test is not valid. The corresponding expression can be used to test the significance of b_u.

In a 4-point assay there are 3 degrees of freedom between doses and two of these are used to calculate slopes. The third may be used to calculate T_1, which is the distance on the zero ordinate between the points where the two uncorrected regression lines cut it. The significance of T_1 can be tested by dividing it by its standard deviation, and consulting a table of t. If it is significant the interpretation of the results by the calculation of slope ratios is not valid. If T_1 is not significant, appropriate corrections are applied to the estimates of the slopes by the formulae given in the table, in which the correction is represented by the last term.

In a 5-point assay there are two degrees of freedom available for tests of validity. They may be used to test the curvature of the two curves separately

TABLE 6

Balanced slope-ratio assays

Number of observations in each dose group ..... n
Variance of effect $= S(d^2)/S(n-1)$ ..... s^2
Variance of each mean effect $= s^2/n$ V
Slopes (Unknown and Standard) ..... b_u and b_s
Variance of slope ..... $V(b)$
Index of significance of $b_s = t^2V(b)/b_s^2$ ..... g
Estimate of potency ratio $U/S = b_u/b_s$ ..... R

	UNKNOWN	STANDARD
Mean effect. Dose 0	Z	
½	U_1	S_1
1	U_2	S_2
3-point assay		
Slope, b	$U_2 - Z$	$S_2 - Z$
Variance of slope, $V(b)$	$2V$	$2V$
Fiducial limits of R	$R + \frac{g}{1-g}\left(R - \frac{1}{2}\right) \pm \frac{t}{1-g}\sqrt{\left(R^2 - R + 1 - \frac{3g}{4}\right)2V/b_s^2}$	
4-point assay		
$T_1 \pm \sqrt{V(T_1)}$	$(2S_1 - S_2 - 2U_1 + U_2) \pm \sqrt{10V}$	
Slope, b	$2(U_2 - U_1 - 3T_1/10)$	$2(S_2 - S_1 + 3T_1/10)$
Variance of slope, $V(b)$	$22V/5$	$22V/5$
Fiducial limits of R	$R + \frac{g}{1-g}\left(R - \frac{9}{11}\right) \pm \frac{t}{1-g}\sqrt{\left(11R^2 - 18R + 11 - \frac{40g}{11}\right)2V/5b_s^2}$	
5-point assay		
Index of curvature	$L_u = Z + U_2 - 2U_1$	$L_s = Z + S_2 - 2S_1$
Variance of L_u or L_s	$6V$	$6V$
$T_1 \pm \sqrt{V(T_1)}$	$L_u - L_s$	$\pm\sqrt{10V}$
$T_2 \pm \sqrt{V(T_2)}$	$L_u + L_s$	$\pm\sqrt{14V}$
Slope, b	$U_2 - Z - T_1/10 + T_2/14$	$S_2 - Z + T_1/10 + T_2/14$
Variance of slope, $V(b)$	$64V/35$	$64V/35$
Fiducial limits of R	$R + \frac{g}{1-g}\left(R - \frac{9}{16}\right) \pm \frac{t}{1-g}\sqrt{\left(8R^2 - 9R + 8 - \frac{175g}{32}\right)8V/35b_s^2}$	

by comparing L_u and L_s with their standard deviations (201). On the other hand, as Finney (82) has shown, they may be used to calculate T_1 and T_2 which correspond to what he calls "Intersection" and "Blanks", respectively. T_1 has the same meaning as in a 4-point assay; if it is significant, the assay is not valid. If T_2 is significant, the effect of zero dose is different from the meeting point of the two regression lines. If this is so, it may still be possible to use the results as a 4-point assay, neglecting the observed effect of zero dose altogether. If neither T_1 nor T_2 is significant, appropriate corrections are applied to the estimates of the slopes by the formulae given in the table.

These methods of calculation are likely to be more widely applied than they have been so far. They can be used whenever a sufficient length of straight line is obtained by plotting the effect against an arithmetic scale of doses, provided that the other general conditions outlined on p. 95 are fulfilled. They are especially suitable when the dose is near the threshold and the data include measurements on control animals which receive no drug at all. In such experiments a logarithmic scale of doses is inconvenient since the effect corresponding to zero dose cannot be directly plotted and can only appear as an asymptote. When an arithmetic scale of doses is used these effects can be easily plotted and the result can be calculated from the slope ratio, if the dose effect lines are straight at their lower ends. This method of plotting has been successfully used to interpret the effects of vitamin D and phosphates on rickets (159), war gases on the reciprocal of the survival time (41), thyrotrophic hormone on thyroid weight (19), ACTH on adrenal weight (14) and androgens on comb length (64).

Variable results have been obtained with gonadotrophins (25, 170). In Levin and Tyndale's data (141) the lower end of the dose-effect line appears to be straight with the ovary but not with the uterus. When the dose is increased sufficiently the effect on uterus weight reaches a maximum and then declines (69).

Any slope-ratio assay can be treated as a parallel-line assay if the mean effect of zero dose is subtracted from all the other readings, and log Y is plotted against log X. This will produce parallel lines if the initial lines are straight and may do so when they are not (201), but does not make the best possible use of the control readings, except when these are made on each animal. In this case it is best to correct each result for the control reading as described above (p. 106).

When the dose-effect curves are not straight they may sometimes be straightened either by pharmacological or by mathematical means. Wood (201) has shown that a small curvature at the lower end of the lines can sometimes be avoided in microbiological assays if a small amount of the factor being studied is added to all the tubes and neglected in the calculations. The same sort of thing might perhaps be done in other types of assay. Dose-effect curves can also be straightened by the use of metameters. It may be convenient in some cases to use dose metameters of the form X^i, where the exact value of i depends on the curvature, being greater than one when the curve is concave upwards, and less when it is convex. If the potency ratio is R this metameter should lead to a slope ratio of R^i.

Effect-metameters can also be used, but in this case it is not so easy to find a formula giving the same status to the effects of zero dose as to the effects of other doses. Emmens (67) has suggested the use of the logistic transformation in interpreting assays depending on the effects of hormones on the weights of organs. He takes Y as the total weight of the organ rather than any change in this weight and uses an arithmetic scale of doses. He believes that the whole curve connecting Y and X is S-shaped and symmetrical, but only the upper part of it can actually be observed, since the curve cuts the axis $X = 0$ at a point corresponding to the weight of the organs when no hormone is given. Although he uses the same transformation as Berkson (23) he makes quite different assumptions about the shape of the curve because he uses an arithmetic scale of doses. Probits could presumably also be used to interpret these curves, but in either case it is necessary to estimate the maximum response and this is not always easy.

XII. ROUTINE TESTS AND COLLABORATIVE ASSAYS

When the same test is repeated frequently it is wise to keep a control chart in which the slope of the curve (b) and the variance of the effect (s^2) are plotted against time (36). Such charts may contain lines enclosing the area within which the result may be expected to vary by chance and an individual result lying outside this area is regarded as suspect. Bliss and Cattell (38) discuss various examples. Knudsen (133) plotted b and s_b. Loraine (146) plotted the actual mean responses.

Gridgeman (111) discussed the results of nearly 500 assays of vitamin A against carotene using (2 and 2) doses. The calculations were all based on intralitter differences since this restriction was found to halve the number of rats used for any given accuracy. The slope of the log-dose-effect curve showed marked seasonal variation. Since the estimate of the slope that can be calculated from the internal evidence of each assay has a fairly large error, Gridgeman recommended that in each assay, b should be calculated from the results of the last 3 assays.

Jones (131) discussed over 100 assays of vitamin D using (3 and 2) and (3 and 3) doses. In 5 out of 6 groups of assays ordinary variance analysis could not be used because the variance (s^2) was not homogeneous. This fact deterred Jones from using the evidence of previous assays in the interpretation of later ones. In the other group of assays there was no significant variation of s or b; the log-dose-effect lines were not significantly curved and even the sensitivity of the animals did not vary significantly. The elimination of the effect of interlitter variance decreased the variance by only about 20 per cent. In such circumstances it is probably not worth the extra work in the laboratory and the sacrifice of degrees of freedom which it involves.

The practical application of methods similar to those described here are illustrated in the results of various collaborative assays organized by official bodies with the object of establishing standard preparations. References to a number of such investigations carried out on behalf of the League of Nations are given by Gautier (104) (cf. especially 66, 149).

Much work has also been done for the British Pharmacopoeia (43, 123), and the United States Pharmacopoeia (32, 152).

REFERENCES

1. Abbott, W. S.: A method of computing the effectiveness of an insecticide. J. Econ. Entomol., 18: 265–267, 1925.
2. Albert, A., and Berkson, J.: A clinical bioassay for chorionic gonadotrophin. J. Clin. Endocrinol., 11: 805–820, 1951.
3. Armitage, P.: Bibliography of Biological Assay 1943–1947 (100 refs.). J. Roy. Statist. Soc., 114: 530–533. 1951.
4. Armitage, P., and Allen, I.: Methods of estimating the LD50 in quantal response data. J. Hyg., 48: 298–322, 1950.
5. Arrhenius, S., and Madsen, T.: Toxines et antitoxines: le poison diphtérique. Oversigt danske Vidensk Selsk. Fork., 4: 269–305, 1904.
6. Bacharach, A. L.: Biological assay and chemical analysis. Inaugural address to Biological Methods Group of the Society of Public Analysts. Analyst, 70: 394–403, 1945.
7. Bacharach, A. L., and Chance, M. R. A.: Interlitter variability as a source of error in gonadotrophin assay. Quart. J. Pharm. and Pharmacol., 18: 10–14, 1945.
8. Bacharach, A. L., Coates, M. E., and Middleton, T. R.: A biological test for vitamin P activity. Biochem. J., 36: 407–412, 1942.
9. Bacharach, A. L., and Cuthbertson, W. F. J.: The cup-plate method of microbiological assay, with special reference to riboflavine and aneurine. Analyst, 73: 334–340, 1948.
10. Bain, W. A.: Comparison of anthisan (mepyramine maleate) and phenergan as histamine antagonists. Lancet, 257: 47–52, 1949.
11. Bain, W. A.: The evaluation of drugs in man with special reference to antihistaminics. Analyst, 76: 573–579, 1951.
12. Baker, G. A.: Linear regression when the standard deviations of arrays are not all equal. J. Am. Stat. Assoc., 36: 500–506, 1941.
13. Basil, B., Somers, G. F., and Woollett, E. A.: Measurement of thyroid activity by the mouse anoxia method. Brit. J. Pharmacol., 5: 315–322, 1950.
14. Bates, R. W., Riddle, O., and Miller, R. A.: Preparation of adrenotropic extracts and their assay on two-day chicks. Endocrinology, 27: 781–792, 1940.
15. Beer, E. J. de: The calculation of biological assay results by graphic methods. The all-or-none type of response. J. Pharmacol. and Exper. Therap., 85: 1–13, 1945.
16. Beer, E. J. de, and Tullar, P. E.: The biological estimation of ergometrine. J. Pharmacol. and Exper. Therap., 71: 256–260, 1940.
17. Behrens, B.: Zur auswertung der digitalisblätter im froschversuch. Arch. Exp. Path. Pharmakol., 140: 237–256, 1929.
18. Behrens, B. and Kärber, G.: Wie sind Reihenversuche für biologische Auswertungen am zweckmässigsten anzuordnen. Arch. f. exper. Path. u. Pharmakol., 177: 379–388, 1934.
19. Bergman, A. J., and Turner, C. W.: A comparison of the guineapig and chick thyroid in the assay of thyrotropic hormone. Endocrinology, 24: 656–664, 1939.
20. Berkson, J.: Application of the logistic function to bio-assay. J. Am. Stat. Assoc., 39: 357–365, 1944.
21. Berkson, J.: Minimum χ^2 and maximum likelihood solution in terms of a linear transform, with particular reference to bioassay. J. Am. Stat. Assoc., 44: 273–278, 1949.
22. Berkson, J.: Some observations with respect to the error of bio-assay. Biometrics, 6: 432–434, 1950.
23. Berkson, J.: Why I prefer logits to probits. Biometrics, 7: 327–339, 1951.
24. Berkson, J.: Relative precision of minimum chi-square and maximum likelihood estimates of regression coefficients. Proceedings of the second Berkeley symposium on mathematical statistics and probability. Univ. California Press, 1951.
25. Bischoff, F.: Ovarian weight increase as an objective measure for assay of prolan. Endocrinology, 30: 667–670, 1942.
26. Bliss, C. I.: The calculation of the dosage-mortality curve. Ann. Appl. Biol., 22: 134–167, 1935.
27. Bliss, C. I.: The comparison of dosage-mortality data. Ann. Appl. Biol., 22: 307–333, 1935.
28. Bliss, C. I.: The calculation of time-mortality data. Ann. Appl. Biol., 24: 815–852, 1937.
29. Bliss, C. I.: The determination of the dosage-mortality curve from small numbers. Quart. J. Pharm. and Pharmacol., 11: 192–216, 1938.
30. Bliss, C. I.: Factorial design and covariance in the biological assay of vitamin D. J. Am. Stat. Assoc., 35: 498–506, 1940.
31. Bliss, C. I.: The relation between exposure time, concentration and toxicity in experiments on insecticides. Ann. Entomol. Soc. Am., 33: 721–766, 1940.
32. Bliss, C. I.: The USP collaborative cat assays for digitalis. J. Am. Pharm. A. (Scient. Ed.), 33: 225–266, 1944.
33. Bliss, C. I.: A simplified calculation of the potency of penicillin and other drugs assayed biologically with a graded dose. J. Am. Stat. Assoc., 39: 479–487, 1944.
34. Bliss, C. I.: Confidence limits for biological assays. Biometrics Bull., 1: 57–65, 1945.
35. Bliss, C. I.: An experimental design for slope-ratio assays. Ann. Math. Stat., 17: 232–237, 1946.
36. Bliss, C. I.: Statistical methods in vitamin research. In Györgyi, P.: Vitamin methods, 2: 445–610, Academic Press Inc., New York, 1951.

37. Bliss, C. I., and Allmark, M. G.: The digitalis cat assay in relation to rate of injection. J. Pharmacol. and Exper. Therap., 81: 378–389, 1944.
38. Bliss, C. I., and Cattell, McK.: Biological assay. Ann. Rev. Physiol., 5: 479–539, 1943.
39. Bliss, C. I., and Marks, H. P.: The biological assay of insulin. Some general considerations directed to increasing the precision of the curve relating dosage and graded response. Quart. J. Pharm. and Pharmacol., 12: 82–110, 1939.
40. Bliss, C. I., and Marks, H. P.: The biological assay of insulin. II. The estimation of drug potency from a graded response. Quart. J. Pharm. and Pharmacol., 12: 182–205, 1939.
41. Box, G. E. P., and Cullumbine, H.: The relationship between survival time and dosage with certain toxic agents. Brit. J. Pharmacol., 2: 27–37, 1947.
42. Boyland, E., McDonald, F. F., and Rumens, M. J.: Variation in the toxicity of phosgene for small animals with the duration of exposure. Brit. J. Pharmacol., 1: 81–89, 1946.
43. British Pharmacopoeia Commission: Report of the subcommittee on the accuracy of biological assays. 1–24, 1936.
44. Bruce, H. M., Parkes, A. S., and Perry, W. L. M.: Assay of ACTH on the thymus of the nestling rat. Lancet, 262: 790–793, 1952.
45. Bryan, W. R.: Quantitative studies of the latent period of tumors induced with subcutaneous injections of the agent of chicken tumor. I. Curve relating dosage of agent and chicken response. J. Nat. Cancer Res. Inst., 6: 225–237, 1946.
46. Burn, J. H., Finney, D. J., and Goodwin, L. G.: Biological standardization. Oxford University Press. London. 1950, 2nd edition.
47. Calloway, N. O., McCormack, R. M., and Singh, N. P.: Studies on the chemistry of the melanorphore hormone of the pituitary gland. I. Critical evaluation of assay methods. Endocrinology, 30: 423–429, 1942.
48. Chance, M. R. A.: Aggregation as a factor influencing the toxicity of sympathomimetic amines in mice. J. Pharmacol. and Exper. Therap., 87: 214–219, 1946.
49. Chance, M. R. A.: Factors influencing the toxicity of sympathomimetic amines to solitary mice. J. Pharmacol. and Exper. Therap., 89: 289–296, 1947.
50. Chang, H. C., and Gaddum, J. H.: Choline esters in tissue extracts. J. Physiol., 79: 255–285, 1933.
51. Churchman, C. W., and Epstein, B.: Tests of increased severity. J. Am. Stat. Assoc., 41: 567–590, 1946.
52. Clark, A. J.: Mode of action of drugs on cells. Edward Arnold, London, 1933.
53. Clark, A. J.: General Pharmacology. Heffters Handbuch experiment. Pharmacol. Ergänzungswerk, 4: 1–228, 1937.
54. Clarke, P. M. Statistical analysis of symmetrical slope-ratio assays at any number of test preparations Abstract. Biometrics, 8: 176–177, 1952.
55. Cornfield, J., and Mantel, N.: Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve. J. Am. Stat. Assoc., 45: 181–210, 1950.
56. Coward, K. H.: The determination of vitamin A in cod-liver oils. Biological examination of the results. Biochem. J., 25: 1104–1106, 1931.
57. Coward, K. H.: The biological standardisation of the vitamins. Baillière, Tindall and Cox, London 1947. 2nd edition.
58. Dale, H. H.: Biological standardization. Analyst, 64: 554–567, 1939.
59. Denton, J. E., and Beecher, H. K.: New analgesics. J. Am. M. A., 141: 1051 and 1146, 1949.
60. Dragstedt, C. A., and Lang, V. F.: Respiratory stimulants in acute cocaine poisoning in rabbits. J. Pharmacol. and Exper. Therap., 32: 215–222, 1927.
61. Durham, F. M., Gaddum, J. H., and Marchal, J. E.: Reports on biological standards. II. Toxicity tests for novarsenobenzene (neosalvarsan). Med. Research Council. Special Report Series, 128: 1–40, 1929.
62. Dutta, N. K., and Macintosh, F. C.: Assay of curare preparations by the rabbit head-drop method. Analyst, 74: 588–592, 1949.
63. Eisenhart, C.: The interpretation of certain regression methods and their use in biological and industrial research. Ann. Math. Statist., 10: 162–186, 1939.
64. Emmens, C. W.: Reports on biological standards: V. Variables affecting the estimation of androgenic and oestrogenic activity. Med. Research Council Special Report Series, 234: 1–71, 1939.
65. Emmens, C. W.: The response of inbred mice to oestrone. J. Endocrinol., 1: 373–377, 1939.
66. Emmens, C. W.: Analysis of the assays of various samples of gonadotrophins. League of Nations. Bulletin of Health Organisation, 8: 862–912, 1939.
67. Emmens, C. W.: The dose/response relation for certain principles of the pituitary gland, and of the serum and urine of pregnancy. J. Endocrinol., 2: 194–225, 1940.
68. Emmens, C. W.: Principles of biological assay. Chapman and Hall, London 1948.
69. Emmens, C. W. (Editor): Hormone assay. Academic Press Inc., New York, 1950.
70. Emmens, C. W.: Standardization. Hormones—a survey of their properties and uses. 111–138. The Pharmaceutical Press, London 1951.
71. Fieller, E. C.: British standard method for the biological assay of Vitamin D_3 by the chick method. British Standards Institution Specification No. 911: 1–20, 1940.
72. Fieller, E. C.: The biological standardization of insulin. J. Roy. Statist. Soc., 7 Suppl. 50, 1941.
73. Fieller, E. C.: A fundamental formula in the statistics of biological assay and some applications. Quart. J. Pharm. and Pharmacol., 17: 117–123, 1944.
74. Fieller, E. C.: Some remarks on the statistical background of bioassay. Analyst, 72: 37–43, 1947.
75. Finney, D. J.: Mathematics of biological assay. Nature, London 153: 284, 1944.

76. FINNEY, D. J.: The application of the probit method to toxicity test data adjusted for mortality in the controls. Ann. appl. Biol., 31: 68–74, 1944.
77. FINNEY, D. J.: The microbiological assay of vitamins: The estimate and its precision. Quart. J. Pharm. and Pharmacol., 18: 77–82, 1945.
78. FINNEY, D. J.: The principles of biological assay. J. Roy. Stat. Soc., 9: 46–76, 1947.
79. FINNEY, D. J.: The adjustment of biological assay results for variation in concomitant observations. J. Hyg., 45: 397–406, 1947.
80. FINNEY, D. J.: The adjustment for a natural response rate in probit analysis. Ann. appl. Biol., 36: 187–195, 1949.
81. FINNEY, D. J.: The choice of a response metameter in bioassay. Biometrics, 5: 261–272, 1949.
82. FINNEY, D. J.: The statistical analysis of slope ratio assays. J. Gen. Microbiol., 5: 223–230, 1951.
82A. FINNEY, D. J.: Biological assay. Brit. Med. Bull., 7: 292–297, 1951.
83. FINNEY, D. J.: Probit analysis, a statistical treatment of the sigmoid response curve. Cambridge Univ. Press, London. 1952. 2nd edition.
84. FINNEY, D. J.: Graphical estimation of relative potency from quantal responses. J. Pharmacol. and Exper. Therap., 104: 440–444, 1952.
84A. FINNEY, D. J.: Statistical method in biological assay. Charles Griffin, London, 1952.
85. FISHER, R. A.: On the dominance ratio. Proc. Roy. Soc. Edin., 42: 321–341, 1921.
86. FISHER, R. A.: The goodness of fit of regression formulae and the distribution of regression coefficients. J. Roy. Statist. Soc., 85: 597–612, 1922.
87. FISHER, R. A.: Statistical methods for research workers. Oliver and Boyd, Edinburgh. 11th edition, 1950.
88. FISHER, R. A., AND YATES, F.: Statistical tables for Biological, Agricultural and Medical Research. Oliver and Boyd, Edinburgh, 1943.
89. GADDUM, J. H.: The action of adrenaline and ergotamine on the uterus of the rabbit. J. Physiol., 61: 141–150, 1926.
90. GADDUM, J. H.: A simple method of measuring surface tension. Proc. Roy. Soc. London B., 109: 114–125, 1931.
91. GADDUM, J. H.: The determination of vitamin A in cod-liver oils. Statistical examination of the results. Biochem. J., 25: 1113–1119, 1931.
92. GADDUM, J. H.: Reports on biological standards III. Methods of biological assay depending on a quantal response. Med. Research Council Special Report Series, 183: 1–46, 1933.
93. GADDUM, J. H.: Discussion on the chemical and physical bases of pharmacological action. Proc. Roy. Soc. London B., 121: 598–601, 1937.
94. GADDUM, J. H.: Contribution to discussion on statistical method applied to biological assays. J. Roy. Statist. A., 4 Suppl.: 49–52, 1937.
95. GADDUM, J. H.: The accuracy of a cross-over test. Quart. J. Pharm. and Pharmacol., 12: 60–65, 1939.
96. GADDUM, J. H.: Pharmacology. Oxford University Press, London. 1st edition, 1940.
97. GADDUM, J. H.: Design of toxicity tests involving comparison with a standard preparation. Quart. J. Pharm. and Pharmacol., 16: 78–86, 1943.
98. GADDUM, J. H.: Lognormal distributions. Nature, London, 156: 463–466, 1945.
99. GADDUM, J. H., AND HETHERINGTON, M.: The activity of thyroid preparations given by the mouth to mice. Quart. J. Pharm. and Pharmacol., 4: 183–194, 1931.
100. GADDUM, J. H., AND LEMBECK, F.: The assay of substances from the adrenal medulla. Brit. J. Pharmacol., 4: 401–408, 1949.
101. GADDUM, J. H., PEART, W. S., AND VOGT, M.: The estimation of adrenaline and allied substances in blood. J. Physiol., 108: 467–481, 1949.
102. GARD, S.: Encephalomyelitis of mice. II. A method for the measurement of virus activity. J. Exper. Med., 72: 69–77, 1940.
103. GARWOOD, F.: The application of maximum likelihood to dosage-mortality curves. Biometrika, 32: 46–58, 1941.
104. GAUTIER, R.: The health organization and biological standardization. League of Nations. Quarterly bulletin of the Health Organization, 4: 497–554, 1935. 12: 1–110, 1945.
105. GERSDORFF, W. A.: A new criterion for the comparison of toxicity with respect to concentration and time. J. Agricult. Res., 50: 881–898, 1935.
106. GLASER, W. R.: Zur theorie der giftwirkungen (Beziehungen zwischen konzentration und einwirkungszeit). Ztschr. gesam. exper. Med., 56: 410–432, 1927.
107. GOLD, H., CATTELL, MCK., OTTO, H. C., KWIT, N. T., AND KRAMER, M. L. A method for the bioassay of digitalis in humans. J. Pharmacol. and Exper. Therap., 75: 196–206, 1942.
108. GRAY, W. H., TREVAN, J. W., BAINBRIDGE, H. W., AND ATTWOOD, A. P.: The ureides of p-Aminophenylstibinic acid. Proc. Roy. Soc. London B, 108: 54–63, 1931.
109. GREEN, A. F., YOUNG, P. A., AND GODFREY, E. I.: A comparison of heat and pressure analgesiometric methods in rats. Brit. J. Pharmacol., 6: 572–585, 1951.
110. GREINER, T., GOLD, H., BLISS, C. I., GLUCK, J., MARSH, R., MATHES, S. B., MODELL, W., OTTO, H., KWIT, N.T., AND WARSHAW, L.: Bioassay of diuretic agents in patients with congestive heart failure. J. Pharmacol. and Exper. Therap., 103: 431–440, 1951.
111. GRIDGEMAN, N. T.: The technique of the biological vitamin A assay. Biochem. J., 37: 127–132, 1943.
112. HANZLIK, P. J.: A study the toxicity of the salicylates based on clinical statistics. J. Am. M. A., 60: 957–962, 1913.
113. HARTLEY, P.: International Biological Standards: Prospect and retrospect. Proc. Roy. Soc. Med., 39: 45–58, 1945.
114. HEALY, M. J. R.: Routine computation of biological assays involving a quantitative response. Biometrics, 5: 330–334, 1949.

115. HEMMINGSEN, A. M., AND MARKS, H. P.: The correlation between the blood sugar fall and initial blood sugar in rabbits injected with insulin. Quart. J. Pharm. and Pharmacol., 5: 245–254, 1932.
116. HEWER, A. J. H., KEELE, C. A., KEELE, K. D., AND NATHAN, P. W.: A clinical method of assessing analgesics. Lancet, 256: 431–434, 1949.
117. HJORT, A. M., BEER, E. J. DE, AND RANDALL, L. O.: Experiences with the biological assay of several sympathicotonic substances including epinephrine. J. Pharmacol. and Exper. Therap., 71: 105–113, 1940.
118. HOLTON, P.: A modification of the method of Dale and Laidlaw for standardization of posterior pituitary extract. Brit. J. Pharmacol., 3: 328–334, 1948.
119. IPSEN, J.: Contribution to the theory of biological standardization. Nyt. Nordisk. Forlag., 1–248, Copenhagen, 1941.
120. IPSEN, J.: A practical method of estimating the mean and standard deviation of truncated normal distributions. Human Biology, 21: 1–16, 1949.
121. IPSEN, J.: Biometric analysis of graded response with incomplete measurements in assays of analgetic drugs. Acta pharmacol. toxicol., 5: 321–346, 1949.
122. IPSEN, J., AND JERNE, N. K.: Graphical evaluation of the distribution of small experimental series. Acta Pathologica, 21: 343–361, 1944.
123. IRWIN, J. O.: Statistical method applied to biological assays. J. Roy. Statist. Soc., Suppl. 4: 1–60, 1937.
124. IRWIN, J. O.: On the calculation of the error of biological assays. J. Hyg., 43: 121–128, 1943.
125. IRWIN, J. O.: A statistical examination of the accuracy of Vitamin A assays. J. Hyg., 43: 291–314, 1944.
126. IRWIN, J. O.: Biological assays with special reference to biological standards. J. Hyg., 48: 215–238, 1950.
127. IRWIN, J. O., AND CHEESEMAN, E. A.: On the maximum likelihood method of determining dosage-response curves and approximations to the median effective dose in cases of a quantal response. J. Roy. Statist. Soc., 6: 174–185, 1939.
128. IRWIN, J. O., AND CHEESEMAN, E. A.: On an approximate method of determining the median effective dose and its error, in the case of a quantal response. J. Hyg., 39: 574–580, 1939.
129. JAQUES, L. B.: A quantitative method of assay for thrombin and prothrombin. J. Physiol., 100: 275–282, 1941.
130. JERNE, N. K., AND WOOD, E. C.: The validity and meaning of the results of biological assays. Biometrics, 5: 273–299, 1949.
131. JONES, J. I. M.: The biological estimation of vitamin D. Quart. J. Pharm. and Pharmacol., 18: 92–108, 1945.
132. KÄRBER, G.: Beitrag zur Kollektiven Behandlung pharmakologischer Reihenversuche. Arch. f. exper. Path. u. Pharmakol., 162: 480–483, 1931.
133. KNUDSEN, L.: Statistics in microbiological assay. Ann. N. Y. Acad. Sci., 52: 889–902, 1950.
134. KNUDSEN, L. F., AND CURTIS, J. M.: The use of the angular transformation in biological assays. J. Am. Stat. Assoc., 42: 282–296, 1947.
135. KNUDSEN, L., AND RANDALL, W. A.: Penicillin assay and its control chart analysis. J. Bact., 50: 187–200, 1945.
136. KNUDSEN, L. F., SMITH, R. B., VOS, B. J., AND MCCLOSKY, W. T.: The biological assay of epinephrine. J. Pharmacol. and Exper. Therap., 86: 339–343, 1946.
137. LANGELAAN, J. W.: On muscle tonus. Brain, 38: 235–380, 1915.
138. LEA, D. E.: The biological assay of carcinogens. Cancer Research, 5: 633–640, 1945.
139. LEHMAN, R. A., AND PAFF, G. H.: A practical technique and design for the assay of digitalis on the embryonic chick heart. J. Pharmacol. and Exper. Therap., 75: 207–218, 1942.
140. LEONG, P. C., AND HARRIS, L. J.: Antineuritic potency of synthetic and natural crystalline vitamin B_1 as determined by the "bradycardia" method. Biochem. J., 31: 672–680, 1937.
141. LEVIN, L., AND TYNDALE, H. H.: The quantitative assay of "follicle stimulating" substances. Endocrinology, 21: 619–628, 1937.
142. LIPSCHITZ, W. L., HADIDIAN, Z., AND KERPCSAR, A.: Bioassay of diuretics. J. Pharmacol. and Exper. Therap., 79: 97–110, 1943.
143. LITCHFIELD, J. T., AND FERTIG, J. W.: On a graphical solution of the dosage-effect curve. Bull. Johns Hopkins Hosp., 69: 276–286, 1941.
144. LITCHFIELD, J. T., AND WILCOXON, F. W.: A simplified method of evaluating dose-effect experiments. J. Pharmacol. and Exper. Therap., 95: 99–113, 1948.
145. LODE, W.: Ein graphisches Verfahren zum Auswerten biologischer Reihenversuche. Medizin und Chemie, 3: 1–9, 1936.
146. LORAINE, J. A.: The estimation of chorionic gonadotrophin in the urine of pregnant women. Addendum on errors by J. H. Gaddum. J. Endocrinol., 6: 319–331, 1950.
147. LORD, E.: The use of range in place of standard deviation in the t-test. Biometrika, 34: 41–67, 1947.
148. MAALØE, O., AND JERNE, N. K.: The standardisation of immunological substances. Animal Rev. Microbiol., 1953.
149. MACRAE, T. F.: The adoption of crystalline vitamin B_1 hydrochloride as the new international standard of vitamin B_1 and comparison of its potency with that of the former standard. League of Nations. Bulletin of Health Organisation, 9: 371–424, 1940.
150. MARKS, H. P.: The biological assay of insulin preparations in comparison with a stable standard. Brit. M. J., 2: 1102, 1925.
151. MILES, A. A.: Biological standards. Brit. Med. Bull., 7: 283–291, 1951.
152. MILLER, L. C.: The U.S.P. collaborative digitalis study using frogs (1939–1941). J. Am. Pharm. A. (Scient. Ed.) 33: 245–266, 1944.
153. MILLER, L. C.: Biological assays involving quantal responses. Ann. N. Y. Acad. Sci., 52: 903–919, 1950.

154. Modell, W., Shane, S. J., Datrit, C., and Gold, H.: Relative potencies of various cinchona alkaloids in patients with auricular fibrillation. Federation Proc., 8: 320–321, 1949.
155. Morgan, R. S.: A diet without caseinogen for use in the determination of vitamin A. Biochem. J., 28: 1178–1192, 1934.
156. Morrell, C. A., Allmark, M. G., and Bachinski, W. M.: On the biological assay of the oxytocic activity of pituitary extract (posterior lobe). J. Pharmacol. and Exper. Therap., 70: 440–449, 1940.
157. Noel, R. H.: The biological assay of epinephrine. J. Pharmacol. and Exper. Therap., 84: 278–283, 1945.
158. du Nouy, L.: Biological time. Methuen and Co., London, 1–180, 1936.
159. O'Brien, B., and Morgaridge, K.: The effect of phosphorus on the biological estimation of vitamin D activity. J. Nutrition, 18: 277–284, 1939.
160. Ostwald, W., and Dernoschek, W.: Über die Beziehungen zwischen Adsorption und Giftigkeit. Kolloid. Z., 6: 297–307, 1910.
161. Paxson, N. F.: Obstetrical anesthesia and analgesia with sodium isoamyl ethyl barbiturate and nitrous oxide-oxygen. Results in obstetrical practice. Anesth. and Analg., 11: 116–122, 1932.
162. Perry, W. L. M.: Reports on biological standards VI. The design of toxicity tests. Med. Research Council Special Report Series, 270: 1–80, 1950.
163. Prentiss, A. M.: Chemicals in war. McGraw-Hill Book Co., Inc., New York, 1937.
164. Pugsley, L. I.: The application of the principles of statistical analysis to the biological assay of hormones. Endocrinology, 39: 161–176, 1946.
165. Reed, L. J., and Muench, H.: A simple method of estimating fifty per cent endpoints. Am. J. Hyg., 27: 493–497, 1938.
166. Reid, E.: Assay of diabetogenic pituitary preparations. J. Endocrinol., 7: 120–142, 1951.
167. Rigg, F. A.: Bibliography 1940–1942. J. Roy. Statist. Soc., 109: 439–442, 1946.
168. Sayers, M. A., Sayers, G., and Woodbury, L. A.: The assay of adrenocorticotrophic hormone by the adrenal ascorbic acid depletion method. Endocrinology, 42: 379–393, 1948.
169. Schild, H. O.: A method of conducting a biological assay on a preparation giving repeated graded responses illustrated by the estimation of histamine. J. Physiol., 101: 115–130, 1942.
170. Sealey, J. L., and Sondern, C. W.: The comparative action of graded doses of international standard gonadotropic substance (chorionic gonadotrophin) on normal infantile male and female rats. Endocrinology, 26: 813–820, 1940.
171. Sherwood, M. B.: Simple formulas for calculating percentage potency in three- and four-dose assay procedures. Science, 106: 153–153, 1947.
172. Smith, A. V., Emmens, C. W., and Parkes, A. S.: Assay of thyroidal activity by a closed vessel technique. J. Endocrinol., 5: 186–206, 1947.
173. Smith, K. W., Marks, H. P., Fieller, E. C., and Broom, W. A.: An extended cross over design and its use in insulin assay. Quart. J. Pharm. and Pharmacol., 17: 108–117, 1944.
174. Smith, R. B., and Vos, B. J.: The biological assay of posterior pituitary solution. J. Pharmacol. and Exper. Therap., 78: 72–78, 1943.
175. Snedecor, G. W.: Statistical methods. The Iowa State College Press. Ames, Iowa, 1946. 4th edition.
176. Spearman, C.: The method of "Right and Wrong Cases" (constant stimuli) without Gauss' formula. Brit. J. Psychol., 2: 227–242, 1908.
177. Squire, J. R.: The relationship between horse dandruff and horse serum antigens in asthma. Clin. Sci., 9: 127–150, 1950.
178. Stevens, W. L.: The truncated normal distribution. Ann. appl. Biol., 24: 847–850, 1937.
179. Tattersfield, F., and Morris, H. M.: An apparatus for testing the toxic values of contact insecticides under controlled conditions. Bull. Entomol. Res., 14: 223–233, 1924.
180. Thompson, R. E.: Biological assay of posterior pituitary. J. Pharmacol. and Exper. Therap., 80: 373–382, 1944.
181. Thompson R. E.: Biological assay of epinephrine. J. Am. Pharm. A. (Scient. Ed.), 34: 265–269, 1945.
182. Thompson, W. R.: Use of moving averages and interpolation to estimate median-effective dose. Bact. Rev., 11: 115–145, 1947.
183. Thompson, W. R.: On the construction of tables for moving average interpolation. Biometrics, 8: 51–54, 1952.
184. Thomson, G. H.: The criterion of goodness of fit of psycho physical curves. Biometrika, 12: 216–230, 1919.
185. Timmerman, W. A.: International biological standardization. J. Pharm. Pharmacol., 3: 65–77, 1951.
186. Trevan, J. W.: The error of determination of toxicity. Proc. Roy. Soc. London B., 101: 483–514, 1927.
187. Trevan, J. W.: A statistical note on the testing of antidysentery sera. J. Path. and Bact., 32: 127–134, 1929.
188. Trevan, J. W.: Accuracy of titration of antipneumococcus serum. J. Path. and Bact., 33: 739–748, 1930.
189. Unna, K. R., Pelikan, E. W. Macfarlane, D. W., Cazort, R. J., Sadove, M. S., Nelson, J. T., and Drucker, A. P.: Evaluation of curarizing drugs on man. I. Potency, duration of action and effects on vital capacity of D-Tubo-Curarine, Dimethyl-D-Tubocurarine and Decamethylene-bis-(Trimethylammonium bromide). J. Pharmacol. and Exper. Therap., 96: 318–329, 1949.
190. Varney, R. F., Linegar, C. R., and Holaday, H. A.: The rabbit "head-drop" method for the biological assay of curare and its alkaloids. Federation Proc., 7: 261–262, 1948.
191. Vogt, M.: The output of cortical hormone by the mammalian suprarenal. J. Physiol., 102: 341–356, 1943.
192. Vos, B. J.: Use of the latent period in the assay of ergonovine on the isolated rabbit uterus. J. Am. Pharm. A. (Scient. Ed.), 32: 138–141, 1943.
193. Waerden, B. L. van der: Wirksamkeits und Konsentrationsbestimmung durch Tierversuche. Arch. f. exper. Path. u. Pharmakol., 195: 389–412, 1940.

194. WEATHERALL, J. A. C. AND WEATHERALL, M.: The effect of dithiols on survival time in rats and mice poisoned with organic arsenicals. Brit. J. Pharmacol., 4: 260–273, 1949.
195. WHITBY, G. S.: Plantation rubber and the testing of rubber. 434–446. Longmans Green and Co., London, 1920.
196. WINDER, C. V.: Misuse of 'deduced ratios' in the estimation of median effective doses. Nature, London, 159: 883, 1947.
197. WINDER, C. V.: Some examples of the use of statistics in pharmacology. Ann. N. Y. Acad. Sci., 52: 838–861, 1950.
198. WINSOR, C. P.: A comparison of certain symmetrical growth curves. J. Washington Acad. Sci., 22: 73–84, 1932.
199 WOOD, E. C.: Mathematics of biological assay. Nature, London, 153: 84–85, 1944. 153: 681–682, 1944.
200. WOOD, E. C.: Calculation of the results of microbiological assays. Nature, London, 155: 632–633, 1945.
201. WOOD, E. C.: The theory of certain analytical procedures, with particular reference to micro-biological assays. Analyst, 71: 1–14, 1946.
202. WOOD, E. C.: The computation of microbiological assays of amino acids and other growth factors. Analyst, 72: 84–90, 1947.
203. WOOD, E. C., AND FINNEY, D. J.: The design and statistical analysis of microbiological assays. Quart. J. Pharm. and Pharmacol., 19: 112–127, 1946.
204. WOOLF, B.: Intrinsic accuracy of biological assays based on a quantitative response. J. Endocrinol., 1953, in press.
205. WRIGHT, H. N.: A simple statistical method for the calculation of mortality percentages in digitalis assays. J. Am. Pharm. A. (Scient. Ed.), 30: 177–180, 1941.
206. YOUNG, P. A.: The duration of cocaine anaesthesia in the rabbit cornea. A statistical examination. Brit. J. Pharmacol., 6: 273–285, 1951.
207. YOUNG, P. A., AND STEWART, G. A.: The distribution of error in mouse insulin assays. J. Pharm. Pharmacol.. 4: 169–180, 1952.
208. YULE, G. U.: On the distribution of deaths with age when causes of death act cumulatively and similar frequency distributions. J. Roy. Statist. Soc., 73: 26–38, 1910.