Chapter 1

INTRODUCTORY

1. Biological Assay

The term *biological assay*, in its widest sense, should be understood to mean the measurement of the potency of any stimulus, physical, chemical or biological, physiological or psychological, by means of the reactions which it produces in living matter. The biological method of measuring the stimulus is adopted either for lack of any alternative, or because an exact physical or chemical measurement of stimulus intensity may need translation into biological units before it can be put to practical use.

Biological assay is most commonly considered as referring to the assessment of the potency of vitamins, hormones, toxicants and drugs of all types by means of the responses produced when doses of these are given to suitable experimental animals. Estimation of the potency of a natural product, such as a drug extracted from plant material, in producing a biological effect of a certain type, is often impossible or impracticable by chemical analysis. Even if the chemical constitution of the material is known or determinable, there may be little knowledge of the magnitude of the effect which the constituents will produce, a difficulty not confined to natural products but occurring also with many manufactured compounds, such as insecticides, which are made to precise chemical specifications yet which are of unknown biological activity. The material must in fact be tested and standardized by methods appropriate to its future use.

For example, vitamin assays may be made in terms of weight changes or other physical measurements observed in rats, the effects of different doses of the preparation to be assayed being compared with the effects of a standard in order to estimate the relative potency of the test preparation and the standard. Insulin may be assayed in terms of the fall in blood sugar in injected rabbits, and digitalis by the mortality amongst injected cats. Again, the potency of insecticides may be assessed by means of

the mortality in batches of treated insects, and that of fungicides by the proportion of treated spores failing to germinate. Another form of assay procedure which is sometimes useful depends on measurement of the time required for the production of a specified effect instead of measurement of the magnitude of the effect produced. In an interesting and informative article, which should be read by all who are seriously concerned with this type of investigation, Bliss and Cattell (1943) have reviewed nearly 300 recently published papers on the theory and practice of biological assay, with especial reference to vitamin, hormone, and drug assay. The texts of Burn (1937) and Coward (1938) may also be consulted, though the statistical methods there advocated do not fully exploit modern developments.

One type of assay which has been found valuable in many different fields, but especially in toxicological studies, is that dependent upon the *quantal*, or all-or-nothing, response. Though quantitative measurement of a response is always to be preferred when available, there are certain responses which permit of no graduation and which can only be expressed as 'occurring' or 'not-occurring'. The most obvious example of this kind of response is death; although workers with insects have often found difficulty in deciding precisely when an insect is dead (Tattersfield *et al.* 1925), in many investigations the only practical interest lies in whether or not a test insect is dead, or perhaps in whether or not it has reached a degree of inactivity such as is thought certain to be followed by early death. In fungicidal investigations, failure of a spore to germinate is a quantal response of similar importance. In studies of drug potency, the response may be the cure of some particular morbid condition, no possibility of partial cure being under consideration. This book is chiefly concerned with assays made by means of quantal responses, though in Chapter 10 some attention is given to quantitative responses. Most of the discussions are presented in terms of tests of the potency of insecticides and fungicides, since it is for these that the methods of analysis were first developed systematically; the same methods, however, are applicable to many other data, both biological and non-biological.

2. Variability of Responses

One feature possessed by all biological assays is the variability in the reaction of the test subjects and the consequent impossibility of reproducing at will the same result in successive trials, however carefully the experimental conditions are controlled. Though similar variability may be encountered in assays based only on purely physical or chemical measurements, it is generally then of far less practical importance. The contrast between the physical approach and the biological may be seen from a consideration of two methods for the estimation of the ratio of two unknown weights. The physical method is to balance each in turn against a set of standardized weights, and to take as the required estimate the ratio of their magnitudes. There may be technical difficulties in carrying out the operations of weighing to very high accuracy, and both the quality of the balance and the competence of the operator are important factors, but for most practical purposes the reproducibility of the results is not called in question; one measurement on each weight will usually suffice to determine the ratio with an accuracy far beyond that obtainable in any biological assay.

The physical assay of the ratio is here so simple that no alternative method is needed. For the sake of the illustration it may be compared with a biological technique, using quantal responses, in which the weights are dropped from a fixed height on to the heads of live rats. Data for the assay are provided by the records of death or survival. That the first weight, at its first trial, killed a rat, while the second weight did not, would not show with any certainty that the first was the heavier, still less would it give any clue to their ratio; the effect would be influenced not only by the weight dropped, but also by the age, sex, size and physical condition of the rat, and other biological and environmental factors (as well as, of course, the shape and elasticity of the weights, which will here be assumed the same for both). If batches of rats, chosen at random from the stock available, were tested with each weight, the proportionate effect of variation in susceptibility from rat to rat would be reduced with increasing size of sample, and

the weights could be compared in terms of the two mortality rates. Variability could be still further controlled, though never entirely eliminated, by using a specially bred strain of rats, and selecting batches homogeneous for sex, age and other relevant factors. When every test is made from the same arbitrary height, this assay cannot discriminate between weights too light to cause any deaths or between weights so heavy as to kill every rat. This difficulty can be overcome by making tests from a series of different heights and obtaining a range of mortalities for each weight. The weights are then compared in terms of equivalent heights, or heights estimated to give the same (say 50 %) mortality. The height scale thus provides a basis for the biological comparison of any number of weights, but, without experimental or theoretical knowledge of the law relating mortality to height and the physical measure of weight, the results of the biological assay cannot be transformed to purely physical terms.

This example has been discussed in some detail, as, in spite of its absurdity, it illustrates the necessity for a careful consideration of variability in any biological assay technique. To some extent the quantal nature of the responses is a complication, but quantitative responses by no means provide an escape from the problem. Equal doses of insulin will not produce equal effects on the blood sugar of different rabbits, or even on the blood sugar of the same rabbit at different times. Consequently, though two insulin preparations could be compared in terms of the magnitudes of the changes in blood sugar produced in two rabbits, only repetition of the tests on several rabbits for each preparation can give an estimate of the relative potency sufficiently precise to be of any practical value.

Biological aspects of, and reasons for, variability in test organisms of many kinds have been discussed by Clark (1933, especially Chapter VI), and his remarks on individual variations in response deserve careful reading. The occurrence of this variability introduces considerations other than those of biology; when there is a large natural variability of response amongst the test subjects, the analysis of numerical data for the estimation of the effects of applied treatments can only be effected satisfactorily with the aid of exact statistical techniques.

3. Statistical Methods

The development of statistical techniques for the analysis of biological data of all types has proceeded with great rapidity in recent years. In many fields of research on biological topics, experimental and observational results can only be used to the best advantage by subjecting them to precise and critical statistical examination. When a programme of biological research involves the collection of numerical data, the problem of interpreting these is almost inevitably one of statistics. The choice is not, as the biologist sometimes imagines, of whether his figures shall be 'statistically analysed' or not, but rather of whether the analysis shall be theoretically sound and able to extract all the relevant information from the material, or inadequate and possibly unsound. Even the simplest and most straightforward averaging of results is essentially a statistical process; the analysis appropriate to any body of data is determined by the inherent properties of those data, not by the whim of the statistician. It is unfortunate, to say the least, that good experimental work should ever be followed by a statistical treatment of the results so unsatisfactory that the conclusions are incomplete, unreliable, or even actively misleading.

The function of the statistician in biological investigations is to supply that critical and objective judgement of numerical material which is a product of his specialized training and experience. An important aspect of his work is co-operation in the planning of an experimental programme so that, taking into account all relevant information already available, it is designed to give results of maximum utility and precision. The assistance of a competent statistician from the beginning of the programme will often substantially increase the value of the results obtained from a given amount of experimental time and labour, in respect of both their scope and their reliability, whereas the conclusions may be much less satisfactory if the statistician is only consulted after the completion of the experimental work.

Nevertheless, the methods of analysis used by the statistician are not esoteric mysteries, but are simply instruments for

discovering the most important features of numerical data. The computational procedures appropriate to many types of data have been so far standardized that they can be applied by a biologist who has some understanding of their purposes, even though he may know little of their theoretical foundations. The blind application of formulae is a danger which should be avoided, for not infrequently the formulae may be used quite inappropriately; on the other hand, the anxiety of many biologists to learn enough of statistical methods in order to be able to analyse their own data without complete dependence on the assistance of a statistician is witnessed by the recent spate of books designed to instruct the non-mathematician in statistical technique.

The statistical treatment of quantal assay data has been much aided by the development of *probit analysis*. This method, which is usually attributed to Gaddum (1933) and Bliss (1934*a*, *b*; 1935*a*, *b*) though it has, in fact, a much longer history (§ 14), has now been widely adopted as the standard method of reducing the data to simple terms.

4. Summary of Contents

This book is written with the intention of introducing the probit method to many who have previously not ventured to use it, and of presenting some of its more recent developments to those who are already familiar with it. In the first few chapters the technique is shown in its simplest form, stripped of all but the essentials. It is hoped that these chapters, at least, will be capable of appreciation and use by many whose knowledge of other branches of statistics is small. Even for this purpose, however, a slight acquaintance with modern statistical thought and terminology is necessary, and although notes on various tests and distributions will be found in the appropriate sections these can do little more than give references and hints on particular applications. The reader is strongly recommended to familiarize himself with the relevant portions of R. A. Fisher's *Statistical Methods for Research Workers* (1944), especially the sections dealing with the normal, t, and χ^2 distributions, and with regression. K. Mather's *Statistical*

Analysis in Biology (1943) provides a valuable introduction for those who find Fisher's book too difficult.

The numerical examples in subsequent chapters have been carefully chosen to illustrate many points of procedure and to show the application of the method to a variety of toxicological data. Though the computational work required is sometimes laborious, it is not as heavy as some accounts have made it appear; in an appendix is given a detailed description of a systematic arrangement of the computations for the simplest type of problem, and this arrangement may easily be extended to suit more complex data. A second appendix gives a brief outline of the mathematical theory of the probit method. The book is completed by a series of tables which lessen considerably the computing time and labour required for probit analysis.

Chapter 2

QUANTAL RESPONSES AND THE DOSAGE-RESPONSE CURVE

5. The Frequency Distribution of Tolerance

In all biological assays there are two components to be considered, the *stimulus* (for example, a vitamin, a drug, a physical force, or a mental test) and the *subject* (for example, an animal, a plant, a piece of tissue, or a single cell). The stimulus is applied to the subject at an intensity specified in units of concentration, weight, time, or other appropriate measure and under environmental conditions as carefully controlled as is practicable, as a result of which a *response* is produced by the subject. Different stimuli are then compared in terms of the magnitudes of the responses they produce, or, more commonly and usefully, in terms of the intensities required to produce equal responses.

When the characteristic response is quantal, its occurrence or non-occurrence will depend upon the intensity of the stimulus applied. For any one subject, under controlled conditions, there will be a certain level of intensity below which the response does not occur and above which the response occurs; in psychology such a value is designated the *threshold* or *limen*, but in pharmacology and toxicology the term *tolerance* seems more appropriate. This tolerance value will vary from one member to another of the population used, frequently between quite wide limits. When the characteristic response is quantitative, the stimulus intensity needed to produce a response of any given magnitude will show similar variation between individuals. In either case, the value for an individual also is likely to vary from one occasion to another as a result of uncontrolled internal or external conditions. Clark (1933, Chapter VI) has discussed the nature of these individual variations in response for many different populations.

For quantal response data it is therefore necessary to consider the distribution of tolerances over the population studied. If the

dose, or intensity of the stimulus, is measured by λ, the distribution of tolerances may be expressed by

$$dP = f(\lambda)\, d\lambda; \tag{2·1}$$

this equation states that a proportion, dP, of the whole population consists of individuals whose tolerances lie between λ and $\lambda + d\lambda$, where $d\lambda$ represents a small interval on the dose scale, and that dP is the length of this interval multiplied by the appropriate value of the *distribution function*,* $f(\lambda)$.

If a dose λ_0 is given to the whole population, all individuals will respond whose tolerances are less than λ_0, and the proportion of these is P, where

$$P = \int_0^{\lambda_0} f(\lambda)\, d\lambda; \tag{2·2}$$

the measure of dose is here assumed to be a quantity which can conceivably range from zero to $+\infty$, response being certain for very high doses so that

$$\int_0^{\infty} f(\lambda)\, d\lambda = 1.$$

The distribution of tolerances, as measured on the natural scale, may be markedly skew, but it is often possible, by a simple transformation of the scale of measurement, to obtain a distribution which is approximately *normal*. 'A variate is said to be normally distributed when it takes all values from $-\infty$ to $+\infty$ with frequencies given by a definite mathematical law, namely, that the logarithm of the frequency at any distance d from the centre of the distribution is less than the logarithm of the frequency at the centre by a quantity proportional to d^2. The distribution is therefore symmetrical, with the greatest frequency at the centre; although the variation is unlimited, the frequency falls off to exceedingly small values at any considerable distance from the centre, since a large negative logarithm corresponds to a very small number' (Fisher, 1944, § 12). In tests of insecticidal sprays, for example, although the distribution of tolerance concentration of the toxic agent is usually far from symmetrical on

* The statement that $f(\lambda)$ is a function of λ means simply that for any given value of λ the value of $f(\lambda)$ is uniquely determined.

account of a few insects with extremely high tolerances providing an extended 'tail' to the distribution (Fig. 1), normalization can often be effected by expressing the tolerances in terms of the logarithms of the concentrations instead of the absolute values (Fig. 2); this transformation is now accepted as standard practice for expressing the results of such trials (cf. Galton, 1879). Various writers (Clark, 1933; Hemmingsen, 1933; Bliss, 1935*a*) have sought an explanation of the normal distribution of log tolerances in the Weber-Fechner law and in adsorption phenomena, particularly as expressed by the Langmuir adsorption law, but these explanations are beyond the scope of this book. The validity and appropriateness of the logarithmic transformation in the analysis of experimental data are not dependent on the truth or falsity of any hypotheses relating to adsorption; use of the log concentration as measuring the dosage in insecticidal trials requires no more justification than that it introduces a simplification into the analysis. There are additional advantages in having a scale on which a given proportionate increase in concentration has the same scale value at all levels of concentration, but other forms of transformation may sometimes be more suitable. Parker-Rhodes (1941, 1942*a*, *b*) has advanced reasons for expecting a normal distribution of some fractional power of the concentration of a fungicide to which suspensions of fungus spores are exposed (see § 45), though this must be only an approximation which holds over a restricted range of concentrations.

It is convenient to take x as representing the intensity of the stimulus on the scale on which the tolerances are normally distributed, and λ as the untransformed value of concentration, time of exposure, or other variate. Thus for much insecticidal work, if λ is the concentration of the toxic agent,

$$x = \log_{10}\lambda, \tag{2·3}$$

and for some fungicides a better transformation may be

$$x = \lambda^i, \tag{2·4}$$

where usually $i \leqslant 1$. The second normalizing transformation tends to the logarithmic as i is decreased to zero. There is no reason why