ance for the breeding of domestic animals, we have started tests on larger domestic animals.

A detailed report of this work will be given elsewhere.

[1] Belling, J., *J. Hered.*, **16**, 461 (1925).
[2] Randolph, L. F., *Proc. U.S. Nat. Acad. Sci.*, **18**, 222 (1932).
[3] Blakeslee, A. F., *C.R. Acad. Sci., Paris*, **205**, 476 (1937); *Gard. Chron.*, **103**, 270 (1938).
[4] Gross, F., *Naturwiss.*, **20**, 962 (1932).
[5] Rostand, J., *C.R. Soc. Biol., Paris*, **113**, 346 (1933).
[6] Griffiths, R. B., *Genetics*, **26**, 69 (1941).
[7] Fankhauser, G., *Proc. Amer. Phil. Soc.*, **79**, 715 (1938); *Biol. Symposia*, **6**, 21 (1942); *Quart. Rev. Biol.*, **20**, 20 (1945).
[8] Häggqvist, G., *Proc. Koninkl. Nederland. Akad. Wetensch.*, **21** (1948).
[9] Pincus, G., and Waddington, C. H., *J. Hered.*, **30**, 515 (1939).
[10] Beatty, R. A., and Fischberg, M., *Nature*, **163**, 807 (1949).

DESIGN AND EVALUATION OF BIOLOGICAL ASSAYS

A QUARTER of a century ago a cumbrous and uncertain tool, to-day a versatile instrument of known precision, biological assay has during its evolution stimulated new biometrical thought as much as pharmacological research; and it would be difficult to say whether this developing use of living organisms in the estimation of biological potency and in quantitative biochemical analysis owes more to practical suggestions from statisticians or to improved statistical devices contributed by biologists not primarily concerned with statistics. Both professions being still in constant need of each other's brains in this field, the Biometric Society (British Region) and the Biological Methods Group of the Society of Public Analysts and Other Analytical Chemists organised a joint meeting in London on March 14, partly to read and discuss papers, and partly as an open forum for questions of mutual interest.

The afternoon session enjoyed the chairmanship of a pioneer of bioassay, Dr. J. W. Trevan, of the Biometric Society. The opening paper, by Mr. N. T. Gridgeman, on "The Graphical Calculation of the Errors of Biological Assays with Graded Responses", described the reduction of the complex expression for the fiducial limits of error, at a given probability level, to

$$I\left[R\,(C-1) \pm \sqrt{(C-1)\,(R^2C+1)}\right],$$

where I is the log dose-interval, R is the response difference (test material minus standard) as a fraction of the dose-interval response difference, and C is Fieller's correction factor. The percentage limits of error are given by 100 times the antilogarithms of this expression, which, incidentally, is applicable only to symmetrical or near-symmetrical assays. The symbol C being a simple function of assay size and response sensitivity, it is possible to construct a nomogram from which C can be read off from n, the total number of animals in the assay, and the statistic λ, the residual standard deviation in terms of the slope of the response/log-dose curve. For a given value of I, a series of error curves can be drawn with C and R as co-ordinates. The chart exhibited, combining the nomogram and the error curves, enables the fiducial limits of a four-point assay to be read off for wide ranges of n, λ and R. A similar chart exists for six-point assays. In presenting the charts, Mr. Gridgeman stressed three advantages in their use: saving of time; reduction of computing mistakes; and easy detection of the factors responsible for unusually narrow or wide limits.

Dr. E. C. Wood then read his paper on "The Estimation of Error in Certain Types of Biological Assays". He began by remarking that it is seldom convenient or economic so to design a graded-response assay that a true estimation of residual error is calculable. A common design employs litters of the same number of animals as there are dosage groups; this necessitates the use of the interaction term, treatments × litters, as the error variance, the true residual term, obtainable only from intra-litter replication of dosages, being inaccessible. However, the occasional availability of big litters permits a doubling-up of treatments within them, and estimation of the true residual variance becomes possible. A collection of relevant data from Dr. Wood's own, and from Mr. A. L. Bacharach's, experimental records had been analysed, and in many instances it was found that the treatments times litters interaction (on which the limits of error of the assay would normally be based) was significantly greater than the residual error. In other instances the total interaction did not bulk large, but one or more of its orthogonal components (for example, slope × litters, parallelism × litters) emerged significantly greater than the residual error. From these observations Dr. Wood concluded that not only is the true residual variance, being sometimes smaller than the interaction term(s), not worth seeking, but also that intra-litter replication is pointless—'extra' animals in litters might as well be discarded, for all the information they can yield.

Correct pooling of the results of non-adjacent and perhaps variously designed assays presents subtle difficulties, some of which were dealt with by Mr. E. C. Fieller in his paper, "The Problem of Combining the Results of Independent Assays". He illustrated one form of the problem with an example from the literature. This comprised three assays of the same material, against the same standard, spaced over a thirteen-month period; all were asymmetric, each was of different size and only one fulfilled on its own account the requirements of biometric validity. Analysis of variance showed, however, that all the embodied estimates of slope, and of residual variance, were statistically homogeneous. Appropriate weighting of the log-doses and of the response differences for each assay yielded valid pooled estimates of the slope and, separately, of the standard minus test-material difference, from which the best estimate of the pooled assay and its limits of error readily follow. It is notable that the worst of the three assays, which, evaluated by itself, yielded limits of error, at the five per cent probability level, from zero to infinity, could yet contribute materially to the pooled assay and help confine its error.

Turning to assays with disparate slopes, Mr. Fieller showed that, in general, the same procedure can apply, for it does not hinge on an assumption of slope homogeneity. If we call the slope B and the standard minus test-material difference Y, the values $B_1, B_2 \ldots B_n$ may differ, but so, proportionally, will the values $Y_1, Y_2 \ldots Y_n$, so that the appropriately weighted $\overline{B}$ and $\overline{Y}$ can be used, just as if $B_1, B_2 \ldots B_n$ were samples of the same population, to obtain an unbiased estimate of the logarithm of the required activity-ratio. Disparity of slope would, of course, be properly reflected in the magnitude of the error of the final estimate.

The evening session of the symposium, with Mr. N. T. Gridgeman, of the Biological Methods Group, in the chair, opened with a discussion of the papers

epitomized above. Mr. D. J. Finney, commenting on the symbolizing by Mr. Gridgeman of the response-difference ratio by R, a letter used elsewhere for the activity ratio, went on to deplore the lack of uniformity in the notation of bioassay statistics generally. The chairman, echoing Mr. Finney's criticisms, expressed the hope that the question of symbol standardization would be settled at an international level, an imperfect but universally accepted notation being better than a perfect one in limited use. Dr. J. I. M. Jones thought that fundamental statistical symbols need standardization, but doubted the practicability of inclusion of derivatory symbols, some of which are short-lived in a world of changing formulæ; the important thing is to collect and define all non-standard symbols in every communication.

Numerous speakers, including Messrs. Fieller, Healy, Bacharach, Finney and Broom, discussed Dr. Wood's paper, the gist of their conclusions being that replications of treatments within litters is understandably of small value; as Mr. M. J. R. Healy put it, the underlying population is litters, and it is desirable to sample this population as widely as possible. Mr. Finney drew an analogy with replicated estimates of a variate n times on m individuals versus single estimates on nm individuals, the latter being far more informative. Dr. Wood concurred, and added as an extension that an assay embracing litters from different laboratories is more valuable than an assay of equal size on litters from a single laboratory.

Concerning Mr. Fieller's paper, Dr. Jones commented on the value of the treatment of the data of the three particular assays in dispelling the superficial inference that the test material had lost potency during the period covered by the assays; the statistical treatment showed that chance would so distribute the results too often to allow weight to be given to apparent time trends. Mr. P. Armitage asked whether the author had compared his method with the common approximate method of weighting the log activity-ratio of each constituent assay with the reciprocal of the variance of that estimate—a method independent of inter-assay differences in variance and slope. Mr. Fieller stated that, in the example given, the weights to be attached to the ill-designed constituent assays were too inaccurate to make the older method usable.

The last part of the meeting was devoted to open discussion of some problems of design and interpretation raised, by previous invitation, by members of the Societies. The problems fell into certain well-defined categories. For example, response curves of unusual position and shape produced questions. Discussion of non-parallel response curves led to the asseveration that the standard and the material under test must not only, in that event, be chemically dissimilar but also must behave differently in the soma; therefore, except in certain special pharmacological cases wherein relative activity is recognized to be a function of dosage-level, the assay must be held invalid. In microbiological assays the occasional peculiarly shaped curve raised difficulties that were argued by Mr. S. A. Price, Mr. J. S. Harrison, and others. It was held to be a biological rather than a statistical problem. So long as the standard and test-material curves are comparable in shape, statistical devices can extract valid assays. The curves can, indeed, as Dr. A. F. Parker-Rhodes mentioned, be mathematically straightened, and Dr. Wood directed attention to his 'doping' technique to linearize the lower parts of response curves.

Designs for assays in which the number of dosage groups exceeds the number of animals per litter were discussed. Mr. Bacharach suggested complete randomization. Mr. Finney thought that partial confounding of interactions as developed in agricultural work provides the best solution. In this connexion Mr. J. V. Smart warned against inadvertent confounding of important interactions, and thought that pilot experiments to assess interactions might be advisable.

Questions were tabled on the influence of uncontrollable variants on responses—for example, the effect of insulin on blood sugar and that of vitamin P on capillary resistance depend on the initial status of the organism. Again, some toxicologists, assuming a relation between body weight and response, dose proportionately instead of so much per animal. The underlying problem is: How are these effects best handled? Messrs. Finney, Wood, Gridgeman and others agreed in advocating initial measurement of the variant, followed by, in all cases (including, within wide limits, toxicity tests), equal doses per animal, and, finally, adjustment where necessary of the responses by covariance analysis to allow for the influence of the uncontrollable variant.

From two sources came queries on the principles governing the selection of animals for test (for example, animals insensitive to rachitogenesis are avoided in vitamin-D assays) and the rejection of data after test (for example, some assayists include a weight loss in growth tests, and others erase the record of that animal, or of its whole litter). A wide-ranging discussion, led by Mr. Bacharach, Dr. Jones, Mr. P. R. Booth, Mr. Finney and Mr. Gridgeman, moved towards the following tentative conclusions: (1) that pre-selection of test animals is justifiable and profitable if the assay is of the purely analytical type, but harbours dangers in potency comparisons of different materials; (2) that, in the absence of observational evidence, no valid criterion exists for the rejection of odd discrepant responses; and (3) that a firm ruling cannot be made on the treatment of weight losses in growth assays, but that their inclusion among results that also contain deaths during the test period involves a logical paradox. As an eminent statistician has declared, some of the points here at issue are moral rather than statistical.

A request for statistical advice on the arrangement of trials involving stepwise dilution initiated a discussion on what was clearly incompletely surveyed territory. Exemplifying the problem, Dr. H. O. J. Collier and Mr. I. F. Hall described tests to compare the efficacy of divers ingested sulphonamides as fæcal bactericides. Pooled fæces from groups of mice so treated were suspended in peptone water and serially diluted in two-times steps. The tubes were autoclaved and infected, and the degrees of bacterial growth after different periods of incubation were recorded, in one experiment, as (−), (− +), (+) and (+ +). In a second experiment, in which the fæces from the individual mice of a group were treated separately, the metameter was the proportion of tubes per group showing growth. Having established that a continuous variate could not be introduced (turbidimetry, for example, would be queered by fæcal particles), the relative merits of the two experimental designs were debated, and it was agreed that the second was the better. Mr. Healy suggested the use of probits, as in toxicity trials, and Mr. Finney spoke on the application of a scoring system to give values amenable to standard statistical treatment.

N. T. Gridgeman