J. Hyg., Camb. (1976), **76**, 147
Printed in Great Britain

Point and interval estimation in the combination of bioassay results

By P. ARMITAGE
Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine

B. M. BENNETT
School of Public Health, University of Hawaii, Honolulu

AND D. J. FINNEY
Department of Statistics, University of Edinburgh

(*Received* 6 *June* 1975)

SUMMARY

A procedure for combining evidence from different biological assays is shown to be equivalent both to generalized least-squares and to maximum-likelihood estimation. By appropriate nesting of hypotheses, the likelihood function can be used to test the agreement between the assays and to obtain probability limits for the combined estimate of potency. The properties of these limits are examined, with particular reference to the situation, unusual but not impossible in practice, in which the values of relative potency that they define consist of several disjoint segments instead of a single interval. The connexion with general theory of estimating linear functional relations is pointed out.

1. INTRODUCTION

The problem of combining evidence from different biological assays, so as to provide a point estimator and confidence intervals for a common potency ratio, is well-known. Finney (1964, chapter 14) discussed several empirical methods that provide point estimators with approximate standard errors, from which approximate limits of error are obtainable. Bennett (1962; a paper henceforth called B) gave a likelihood ratio statistic from which a point estimator and confidence limits could be derived (see also Bennett, 1963*a*).

Armitage (1970; henceforth called A) described a method for obtaining the maximum-likelihood (ML) estimator of the common potency ratio. Subsequently, Armitage & Bennett (1974; henceforth called AB) confirmed that the A and B estimators are identical. The present paper examines the general theory of combining assay results in greater detail; it then considers the particular problem of interval estimation, and finds therein some unusual features. The notation is largely consistent with that in AB, though with some generalizations.

As noted in A, the problem can be regarded as a particular case of the estimation of a linear functional relation between two variables. Some of the results in §3 can

be found in the literature on functional relations, to which references are given below. Other results, such as the use of (4.3) and (4.5) and the discussions in §§5, 6, may be of interest in the wider context of functional relations.

2. SPECIFICATION OF PROBLEMS

Suppose that each of k parallel-line assays compares the same standard and test preparations, the true regression equation in assay i being

$$\left.\begin{aligned} \eta_i &= \alpha_i + \beta_i x, \\ \eta_i' &= \alpha_i' + \beta_i x, \end{aligned}\right\} \tag{2.1}$$

for standard and test respectively, where x is the logarithm of dose. Then the log potency ratio of test relative to standard in assay i is

$$\mu_i = (\alpha_i' - \alpha_i)/\beta_i. \tag{2.2}$$

Suppose that $\bar{x}_i$ and $\bar{x}_i'$ are the mean values of log dose for the two preparations, $\bar{y}_i$ and $\bar{y}_i'$ are the observed mean responses, and B_i the estimated regression coefficient in assay i. Equations (2.1) are estimated by

$$\left.\begin{aligned} Y_i &= \bar{y}_i + B_i(x - \bar{x}_i), \\ Y_i' &= \bar{y}_i' + B_i(x - \bar{x}_i'). \end{aligned}\right\} \tag{2.3}$$

Write

$$D_i = \bar{y}_i' - \bar{y}_i, \tag{2.4}$$

$$\delta_i = E(D_i) \tag{2.5}$$

and

$$z_i = \bar{x}_i - \bar{x}_i', \tag{2.6}$$

the last being a constant determined by the assay design. Then (2.2) can be expressed

$$\mu_i = z_i + \delta_i/\beta_i. \tag{2.7}$$

We shall be concerned particularly with hypotheses that specify equality of the μ_i.

We shall adopt the customary assumption that individual responses in assay i are distributed normally about the appropriate regression lines. We shall further assume the variance of the distribution to be a constant, σ^2, for all the assays, this condition being at the heart of the argument that follows. For assay i, the usual least-squares procedure for fitting parallel regression lines provides unbiased estimators B_i, D_i of β_i, δ_i, and these together with the residual sum of squares are sufficient statistics for β_i, δ_i, σ^2. For the data as a whole, the k pairs (B_i, D_i) and the pooled residual sum of squares about the $2k$ parallel lines are sufficient for the $(2k+1)$ parameters. Most of the following discussion is in terms of the set (B_i, D_i) rather than individual responses.

In general, B_i and D_i are normally distributed about β_i and δ_i, with

$$\left.\begin{aligned} \text{var } (B_i) &= u_i\sigma^2, \\ \text{var } (D_i) &= v_i\sigma^2, \\ \text{cov } (B_i, D_i) &= w_i\sigma^2, \end{aligned}\right\} \tag{2.8}$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

where u_i, v_i, w_i depend solely on the design and number of responses measured for assay i. In the simpler types of assay corresponding to equations (2.1)–(2.8), w_i will be zero; complexities of design such as incomplete blocks or missing observations commonly introduce a covariance, and w_i is introduced here for generality.

Although the emphasis of this paper is on parallel-line assays, essentially the same theory is applicable to slope-ratio assays (Bennett, 1963*b*). For suppose that x is now the absolute dose and, in a notation that is non-standard but convenient here, equations (2.1) are replaced by

$$\left.\begin{aligned} \eta_i &= \alpha_i+\beta_i x, \\ \eta_i' &= \alpha_i+\delta_i x. \end{aligned}\right\} \tag{2.9}$$

Then the relative potency is

$$\mu_i = \delta_i/\beta_i. \tag{2.10}$$

Again under the customary assumptions of normality, least-squares theory leads to statistics B_i, D_i that are sufficient and unbiased estimators of β_i, δ_i for assay i. The constraint that the two regression lines shall meet at $x = 0$ introduces a non-zero covariance between B_i, D_i, and the generality of (2.8) is needed.

Equation (2.7) appears to be more complex than (2.10), in that z_i can be different in each assay. However, if δ_i^* is defined by

$$\delta_i^* = \delta_i+\beta_i z_i \tag{2.11}$$

and estimated by D_i^*, where

$$D_i^* = D_i+B_i z_i, \tag{2.12}$$

then (2.7) can be put in exactly the same form as (2.10). If B_i and D_i are uncorrelated, B_i and D_i^* will have a non-zero covariance. In fact, either of z_i, w_i can be regarded as redundant since a simple transformation allows the problem to be expressed in a form involving only one of them. For practical convenience, both are retained in this paper.

The central problem, then, is that of estimating a ratio of two parameters, where k pairs of unbiased and normally distributed estimators of the two are available, and the covariance matrix for every pair consists of three known multiples of a single variance that itself can be estimated from a pooled sum of squares.

3. THE LIKELIHOOD

The most general hypothesis, H_0, puts no constraints on the parameters and in particular permits all μ_i to be different. Under H_0, the log likelihood is

$$L_0 = C(\sigma^2)-\frac{1}{2\sigma^2}\sum\frac{1}{(u_i v_i-w_i^2)}\{v_i(B_i-\beta_i)^2-2w_i(B_i-\beta_i)(D_i-\delta_i)+u_i(D_i-\delta_i)^2\}, \tag{3.1}$$

where $C(\sigma^2)$ is a function of σ^2 and of individual responses that does not involve the β_i, δ_i and summation (as elsewhere in the paper) is from $i = 1$ to $i = k$. The maximum of L_0 is

$$L_{0,\max} = C_0,$$

occurring when $\hat{\beta} = B_i$, $\hat{\delta}_i = D_i$, and $\sigma^2 = \hat{\sigma}_0^2$ is chosen to maximize $C(\sigma^2)$. The μ_i are estimated by

$$\hat{\mu}_i = z_i+B_i/D_i. \tag{3.2}$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

We now consider two other composite hypotheses involving constraints on the parameters β_i, δ_i:

$$H_{k-1}: \mu_i = \mu \text{ (unspecified)},$$

$$H_k: \mu_i = \mu_0 \text{ (a specified value)}.$$

The subscripts on H indicate the number of constraints imposed; H_{k-1} is a subset of H_0 and H_k is a subset of H_{k-1}. We shall study the maximization of likelihood, first under H_k and then under H_{k-1}.

From (2.7) and under H_k,

$$\delta_i = \beta_i(\mu_0 - z_i). \tag{3.3}$$

The log likelihoood is

$$L_k = C(\sigma^2) - \frac{1}{2\sigma^2}\sum \frac{1}{(u_i v_i - w_i^2)} \times \{v_i(B_i-\beta_i)^2 - 2w_i(B_i-\beta_i)(D_i-\beta_i\xi_i) + u_i(D_i-\beta_i\xi_i)^2\}, \tag{3.4}$$

where ξ_i is written for $(\mu_0 - z_i)$. Algebraic manipulation enables this to be written

$$L_k = C(\sigma^2) - \frac{1}{2\sigma^2}\sum \frac{(B_i\xi_i - D_i)^2}{u_i\xi_i^2 - 2w_i\xi_i + v_i} - \frac{1}{2\sigma^2}\sum \frac{\{\beta_i(u_i\xi_i^2 - 2w_i\xi_i + v_i) - \xi_i(u_iD_i - w_iB_i) - (v_iB_i - w_iD_i)\}^2}{(u_i\xi_i^2 - 2w_i\xi_i + v_i)(u_iv_i - w_i^2)}. \tag{3.5}$$

Only the final summation involves β_i, and the δ_i have been eliminated. Hence L_k is maximized by making all terms in that summation zero. The ML estimates are therefore

$$\left.\begin{aligned} \hat{\beta}_i &= \frac{\xi_i(u_iD_i - w_iB_i) + (v_iB_i - w_iD_i)}{u_i\xi_i^2 - 2w_i\xi_i + v_i}, \\ \hat{\delta}_i &= \hat{\beta}_i\xi_i \end{aligned}\right\} \tag{3.6}$$

and

$$L_{k,\max} = C_k - \frac{1}{2\hat{\sigma}_k^2}\sum \frac{(B_i\xi_i - D_i)^2}{u_i\xi_i^2 - 2w_i\xi_i + v_i}, \tag{3.7}$$

where $\hat{\sigma}_k^2$ is the value of σ^2 maximizing L_k and $C_k = C(\hat{\sigma}_k^2)$.

Maximization under H_k leads easily to explicit formulae for the estimates of β_i, δ_i. Under H_{k-1}, the known μ_0 is replaced by an unknown μ that must also be estimated. Evidently for any μ equations (3.6) determine corresponding maximizing values for β_i, δ_i, and therefore the ML estimates under H_{k-1} can be obtained by maximizing $L_{k,\max}$ with respect to μ. Sprent (1966) followed the same procedure in a generalized least-squares solution to the functional relation problem; the summation in (3.7) is the same as Sprent's (10). Dolby (1972) demonstrated the equivalence of maximum likelihood and generalized least squares in the estimation of a linear functional relation. Replacement of ξ_i by $(\mu - z_i)$ in (3.4) and differentiation with respect to μ gives the equation

$$\sum \frac{w_i\beta_i(B_i-\beta_i) - u_i\beta_i(D_i - \beta_i\mu + \beta_i z_i)}{u_iv_i - w_i^2} = 0$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

as an additional condition on the ML estimates. Hence

$$\hat{\mu} = \Sigma \frac{\hat{\beta}_i\{(u_i D_i - w_i B_i) + \hat{\beta}_i(u_i z_i + w_i)\}}{u_i v_i - w_i^2} \Big/ \Sigma \frac{u_i \hat{\beta}_i^2}{u_i v_i - w_i^2}. \tag{3.8}$$

Iteration with equations (3.6), (3.8) enables $\hat{\mu}$ and the $\hat{\beta}_i$ to be obtained numerically to any desired accuracy. Except for the introduction of the covariance, this is the same as the procedure using equations (1), (2) of paper A.

The maximum log likelihood under H_{k-1} is obtained from (3.7) by putting $\mu_0 = \hat{\mu}$ and choosing $\sigma^2 = \hat{\sigma}^2_{k-1}$ to maximize the expression. Writing for conciseness $\hat{\xi}_i = \hat{\mu} - z_i$,

$$L_{k-1,\max} = C_{k-1} - \frac{1}{2\hat{\sigma}^2_{k-1}} \Sigma \frac{(B_i \hat{\xi}_i - D_i)^2}{u_i \hat{\xi}_i^2 - 2w_i \hat{\xi}_i + v_i}, \tag{3.9}$$

where $C_{k-1} = C(\hat{\sigma}^2_{k-1})$.

4. LIKELIHOOD RATIOS AND RELATED TESTS

The hypothesis H_k is a constrained form of H_{k-1}, which in turn imposes additional constraints on H_0. Although these constraints are non-linear, only H_0 being a truly linear model, general theory of hypotheses nested in this fashion ensures that

$$Z_{k-1} = 2(L_{0,\max} - L_{k-1,\max}) \tag{4.1}$$

can be used to test H_{k-1} within the wider family H_0 by regarding it as asymptotically distributed as $\chi^2_{[k-1]}$. Similarly, H_k can be tested within the wider family H_{k-1} by regarding

$$Z_1 = 2(L_{k-1,\max} - L_{k,\max}) \tag{4.2}$$

as a $\chi^2_{[1]}$. Thus (4.1) and (4.2) enable the heterogeneity of the μ_i and the deviation of a common μ from a specified μ_0 to be tested.

The algebraic forms are complicated by the fact that the maximizations involve three different estimators of σ^2. Asymptotically, this can be neglected, and formal proof along standard lines confirms the more heuristic conclusion from (3.8) that

$$Z^*_{k-1} = \frac{1}{\hat{\sigma}^2} \Sigma \frac{(B_i \hat{\xi}_i - D_i)^2}{u_i \hat{\xi}_i^2 - 2w_i \hat{\xi}_i + v_i} \tag{4.3}$$

may be used in place of Z_{k-1}; here $\hat{\sigma}^2$ can be any efficient estimator of σ^2, for example the residual mean square from the unconstrained model, $\hat{\sigma}^2_0$. When every $w_i = 0$, substitutions from (3.6) allow the last expression to be written

$$Z^*_{k-1} = \frac{1}{\hat{\sigma}^2} \Sigma \frac{(B_i - \hat{\beta}_i)^2}{u_i} \left(1 + \frac{v_i}{u_i \hat{\xi}_i^2}\right), \tag{4.4}$$

a formula given by Armitage, Bailey, Petrie, Annable & Stack-Dunne (1974, eq. (11)) for the case σ^2 known. Similarly Z_1 may be replaced by

$$Z^*_1 = \frac{1}{\hat{\sigma}^2} \Sigma \left\{ \frac{(B_i \xi_i - D_i)^2}{u_i \xi_i^2 - 2w_i \xi_i + v_i} - \frac{(B_i \hat{\xi}_i - D_i)^2}{u_i \hat{\xi}_i^2 - 2w_i \hat{\xi}_i + v_i} \right\}. \tag{4.5}$$

If $\hat{\sigma}^2$ is an ordinary quadratic estimator of σ^2 with f degrees of freedom, as it is when equated to $\hat{\sigma}^2_0$, Z^*_{k-1} can more plausibly be asserted to be distributed as

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

$(k-1)\,F(k-1, f)$, and we conjecture that the distribution of Z_{k-1} is better approximated by this than by $\chi^2_{[k-1]}$. The distinction disappears when f becomes large. Similarly, the distribution of $F(1, f)$ may be a better approximation than that of $\chi^2_{[1]}$ for Z_1^* and Z_1.

Another approach to an approximate test statistic is to obtain the asymptotic variance of $\hat{\mu}$ by evaluating second derivatives of L_k with respect to μ, β_1, β_2, ..., β_k and σ^2, and inverting the matrix. This leads to

$$\text{var}\,(\hat{\mu}) = \sigma^2 \left\{ \Sigma \frac{\beta_i^2}{u_i \xi_i^2 - 2w_i \xi_i + v_i} \right\}^{-1}, \tag{4.6}$$

a result obtained in A for the particular case of $w_i = 0$. Hence an asymptotic $\chi^2_{[1]}$ statistic is

$$Z_1^{**} = (\hat{\mu} - \mu_0)^2 / \text{var}\,(\hat{\mu}), \tag{4.7}$$

where $\hat{\theta}_i$, $\hat{\beta}_i$, $\hat{\sigma}^2$ are inserted for θ_i, β_i, σ^2 in (4.6) in order to permit evaluation of var $(\hat{\mu})$. General theory shows Z_1^* and Z_1^{**} to be asymptotically equivalent, and this can be verified directly by expanding Z_1^* as a power series in $(\hat{\mu} - \mu_0)$.

In B, the suggestion was made that H_k could be tested by regarding the second term of (3.7), which is asymptotically equivalent to $Z_1 + Z_{k-1}$, as being distributed as $kF(k, f)$, or approximately $\chi^2_{[k]}$. Similar proposals have been made elsewhere (Brown, 1957; Sprent, 1966, 1969; Villegas, 1964). General considerations indicate that a test based on Z_1 or Z_1^* using $F(1, f)$ or $\chi^2_{[1]}$ will be more powerful in detecting deviation of μ from the μ_0 specified by H_k. Beale (1966) made a similar comment in relation to interval estimation.

5. THE ESTIMATION OF POTENCY

Consider now the function

$$J(\mu) = \Sigma \frac{\{B_i(\mu - z_i) - D_i\}^2}{u_i(\mu - z_i)^2 - 2w_i(\mu - z_i) + v_i}. \tag{5.1}$$

It can be regarded in three ways:

(i) By reference to (3.7), (3.9), minimization of $J(\mu)$ can be seen as equivalent to maximization of the likelihood if σ^2 is known, and as part of the condition for maximizing the likelihood even if σ^2 is unknown.

(ii) Under the hypothesis H_{k-1}, $J(\mu)/\sigma^2$ is distributed as χ^2 with k D.F. If an estimate of σ^2, say $\hat{\sigma}^2$ with f degrees of freedom, is available from a statistically independent source, $J(\mu)/k\hat{\sigma}^2$ is distributed as $F(k, f)$.

(iii) Since $E[B_i(\mu - z_i) - D_i] = 0$, $J(\mu)$ can be regarded as a sum of squares of deviations of the quantities $\{B_i(\mu - z_i) - D_i\}$ from their expectations, each square being weighted in inverse proportion to its variance.

Thus determination of $\hat{\mu}$ to minimize $J(\mu)$ can be seen to be a maximum-likelihood, minimum χ^2, and non-linear least-squares estimation procedure.

Evidently as μ becomes large, positive or negative, $J(\mu)$ approaches

$$J_\infty = \Sigma \frac{B_i^2}{u_i}. \tag{5.2}$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

A graph of $J(\mu)$ against μ will show J_∞ as a horizontal asymptote at both extremes. Note that necessarily $u_i > 0$, and that the question of combining assay results would not arise unless some $B_i \neq 0$, so that the asymptotic value exceeds zero. By expansion in inverse powers of μ,

$$J(\mu) = \Sigma \frac{B_i^2}{u_i} - \frac{2}{\mu} \Sigma \left(\frac{B_i D_i}{u_i} - \frac{w_i B_i^2}{u_i^2} \right) + \ldots . \tag{5.3}$$

Hence in general the asymptote will be approached from opposite sides at the positive and negative extremes of μ. A necessary condition for the approach to be from one side only is that

$$\Sigma \left(\frac{B_i D_i}{u_i} - \frac{w_i B_i^2}{u_i^2} \right) = 0. \tag{5.4}$$

This condition does not represent any obviously interesting special case, though undoubtedly it can be satisfied. The term in μ^{-3} in (5.3) would be more complicated and is not shown here; certainly equation (5.4) does not oblige it also to vanish. Consequently the usual situation is that $J(\mu)$ approaches its asymptote from opposite sides, but approach from the same side at both ends may occur for certain non-trivial conjunctions of values of the B_i, D_i, u_i, v_i, w_i.

Fig. 1 is a sketch of a typical curve for data in which all estimates of μ from the single assays agree well so that H_{k-1} is reasonable. The curve will take either this form or its mirror image according to the sign of the expression on the left of equation (5.4). It is qualitatively similar to the well-known graphical representation of Fieller's theorem, except that the minimum of $J(\mu)$ is greater than zero unless the k terms of $J(\mu)$ differ only by known numerical factors. Of course if $k = 1$ the curve is exactly Fieller's.

Before variants of the curve are considered, the interpretation of the simple form requires comment. The minimum is $J(\hat{\mu})$, where $\hat{\mu}$ is the estimate of μ under H_{k-1} obtained by equation (3.7). If σ^2 were known,

$$J(\hat{\mu})/\sigma^2 \tag{5.5}$$

would be asymptotically distributed as $\chi^2_{[k-1]}$ and could test H_{k-1}: evidently if the minimum is large the hypothesis that the separate assays share a common μ is implausible. This test is identical with the Z_{k-1} test of equation (4.1) for the special case of a known σ^2, since the minimum of $J(\mu)$ under H_0 is zero. In practice, if several assays are being combined with the assumption that σ^2 is the same for all, the number of degrees of freedom for estimating σ^2 will usually be sufficiently large to permit substitution of $\hat{\sigma}^2$ for σ^2 without much harm. One may expect that, if $\hat{\sigma}^2$ is taken from a composite intra-dose sum of squares with f degrees of freedom, a better test would be to regard $J(\hat{\mu})/(k-1)\hat{\sigma}^2$ as having the $F(k-1,f)$ distribution, but general theory permits the $\chi^2_{[k-1]}$ test to be used asymptotically just as for Z_{k-1}.

Probability limits can be assigned to μ by similar argument from the Z_1 test of equation (4.2). When σ^2 is known, the acceptable values of μ at probability P are those for which

$$[J(\mu) - J(\hat{\mu})]/\sigma^2 \leqslant \chi^2_{[1]}, \tag{5.6}$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

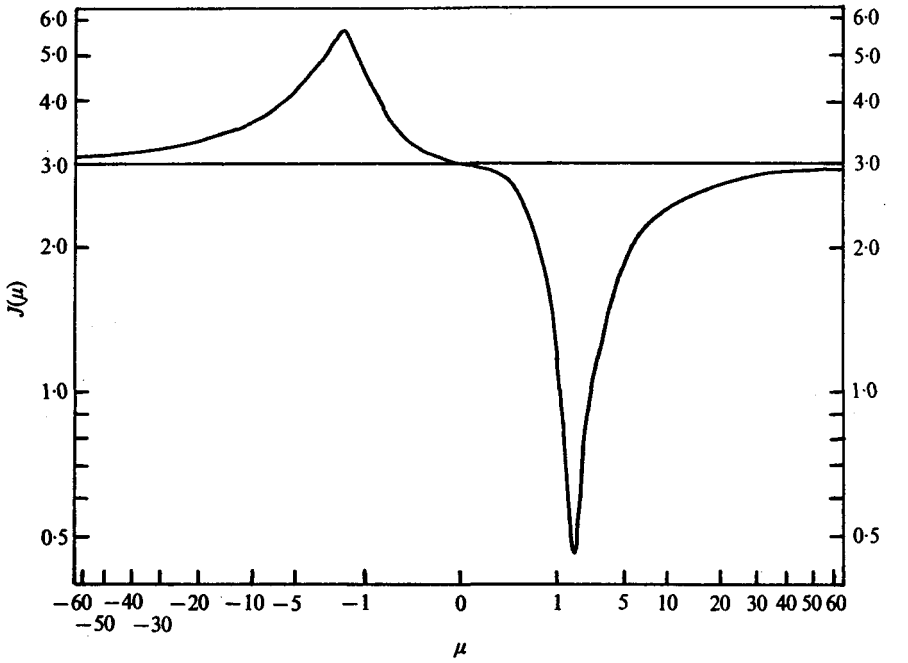

Fig. 1. The function $J(\mu)$ for $k = 3$ and

i	B_i	D_i	z_i	u_i	v_i	w_i
1	1	1	0	1	1	0
2	1	2	0	1	2	−1
3	1	0	0	1	2	1

[In order to show extremes more clearly, the scale of μ is drawn as linear in $\mu^{\frac{1}{3}}$ and that of $J(\mu)$ as linear in $\log J$.]

where the value of χ^2 is read for probability P. Replacement of σ^2 by $\hat{\sigma}^2$ leaves the test asymptotically unchanged, though substitution of $F(1, f)$ for $\chi^2_{[1]}$ may be an improvement (Beale, 1960). In accordance with standard statistical practice, this procedure is appropriate when H_{k-1} is known to be true and limits are to be stated irrespective of whether or not the test of (5.5) supports the hypothesis. If the truth of H_{k-1} is uncertain, either because of doubts whether the k assays are all concerned with the same potency or because one or more assays have been contaminated in some way, the test of (5.5) takes the role of a validity test. Limits calculated by use of (5.6) are conditional on H_{k-1}, but judgement on whether to regard the limits as meaningful, rather than rigid reliance upon a test of significance, seems desirable.

For $k = 1$, equation (5.6) or its modification with a value of F (or t^2) in place of χ^2 leads to the quadratic inequality that underlies Fieller's theorem. The values of μ satisfying the inequality form a single finite interval if at the chosen probability level

$$J_\infty/\sigma^2 > \chi^2_{[1]}, \tag{5.7}$$

and form an infinite interval from which a finite segment may be excluded if this inequality is reversed. The situation is clear from consideration of placing a horizontal line at height $\sigma^2\chi^2_{[1]}$ on Fig. 1, remembering that $J(\hat{\mu}) = 0$ when $k = 1$. If $\sigma^2\chi^2_{[1]}$ exceeds the maximum of $J(\mu)$, the whole range of μ from $-\infty$ to $+\infty$ must

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

be regarded as not rejected at the stated probability. For larger k, the argument remains the same except that the inequality corresponding to (5.7) is now

$$[J_\infty - J(\hat{\mu})]/\sigma^2 > \chi^2_{[1]}, \tag{5.8}$$

and the height of the horizontal inserted in Fig. 1 must be measured from $J(\hat{\mu})$.

If the ratios $u_i : v_i : w_i$ are constant for all i, as would happen if all assays were the same in design but might be differently replicated, the curve will certainly be of the form of Fig. 1. Trials suggest that even quite large discrepancies in these ratios still leave the curve with one minimum and one maximum. As will be illustrated in § 6, however, the function $J(\mu)$ can have two or more minima, which in general will not be equal. Indeed, since a single term of $J(\mu)$, corresponding to one assay, itself produces a curve with one minimum and one maximum and a horizontal asymptote at each extreme, presumably the general expression can be so contrived as to have k minima and k maxima. All that is necessary is to sum terms so displaced relative to one another that the minimum and maximum for each taken alone occur far out towards the asymptote of every other one. A formal proof could doubtless be given, but the nature of it should be apparent. Whether a $J(\mu)$ can ever have more than k minima is unknown.† However many they are, the local minima will be separated by maxima that can be of any height: some maxima may rise far above the asymptote, others may be barely discernible separators of adjacent minima. The asymptote remains at J_∞, and the most general situation, k minima (some of which may be above the asymptote) and k maxima, will have an approach to the asymptote from above beyond the outlying maximum. Figs. 2 and 3 illustrate examples for $k = 2, 3$.

In most problems of applied statistics, likelihood functions appear to behave in a simple fashion. Multiple maxima are unusual, except perhaps for trivial special values of parameters, and the function is commonly approximated by a quadratic satisfactorily over a large region. For the present problem in the special case of σ^2 known, $-J(\mu)$ is proportional to the log-likelihood under hypothesis H_{k-1}. As Figs. 2 and 3 indicate, this log-likelihood for a single parameter can manifest an extraordinary diversity of form. Such remarkable behaviour by a univariate likelihood for a relatively simple model has seldom received comment, though Edwards (1972, § 8.2) has an interesting example relating to estimation of the scale factor for a Cauchy distribution (see also Barnett, 1966).

The absolute minimum of $J(\mu)$ still corresponds to $\hat{\mu}$, the maximum likelihood, minimum χ^2, and least-squares estimate of μ. To express precision of this estimate by an asymptotic variance will be most unwise, unless all other minima are at substantially larger values of $J(\mu)$: this is surely clear from consideration of what inference is appropriate if a second minimum has $J(\mu)$ almost as small at a value of μ very different from $\hat{\mu}$. Interval estimates, however, can be ascribed to μ by exactly the same procedure as before. Using a chosen probability level, a horizontal line can be placed on the diagram at a height

$$J(\hat{\mu}) + \sigma^2 \chi^2_{[1]} \quad \text{or} \quad J(\hat{\mu}) + \hat{\sigma}^2 F(1, f).$$

† Dr W. Knight has shown us a simple demonstration that as many as $2k - 1$ minima and $2k - 1$ maxima can occur.

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

The values of μ within the probability set are then those for which the curve lies below the line. The set can consist of several disjoint segments, of which one may include infinity. Segments can coalesce, and for some probabilities the number of segments will be smaller, down even to 1.

Note that the form of $J(\mu)$ is independent of σ^2. Whatever the value of σ^2, for some levels of probability all possible types of intersection with the horizontal will occur; the pattern that corresponds to $P = 0{\cdot}95$ for a small σ^2 may become that for the less interesting $P = 0{\cdot}08$ when σ^2 is large. In practice, if the k assays are reasonably precise and agree sufficiently well to make a combined estimate desirable, it seems unlikely that the more bizarre types of probability set will be found at levels of probability usually considered interesting. Only if the truth of H_{k-1} rested upon firm knowledge outside the current assays could a set of values determined at a probability in the range 0·90–0·99 and found to consist of several disjoint segments be regarded as conveying useful information. More commonly, the occurrence of a set at such a probability that consists of clearly separated segments would correspond with evidence against H_{k-1} sufficiently strong to force *either* rejection of the whole hypothesis *or* recognition that some of the constituent assays are untrustworthy and possibly invalid.

6. TWO ARTIFICIAL EXAMPLES

When an attempt is made to combine assays that are individually precise and in good agreement with one another, the curve for $J(\mu)$ will be qualitatively similar to Fig. 1. Some study of the pathology of the problem is interesting, and two examples will be considered here.

First suppose that $k = 2$, and take the hypothetical scheme of values:

i	B_i	D_i	z_i	u_i	v_i	w_i
1	1	$-d$	0	a	1	0
2	1	d	0	1	a	0

for any $d \geqslant 0$ and $a \geqslant 0$. Evidently

$$\left.\begin{aligned} \hat{\mu}_1 &= -d, \\ \hat{\mu}_2 &= d. \end{aligned}\right\} \tag{6.1}$$

This is not intended to represent a particularly realistic bioassay situation, but it it is not impossible. Moreover, if σ^2 is sufficiently small, the precision of each assay can be assessed in terms of its asymptotic variance, here easily reduced to the form

$$\left.\begin{aligned} \text{var}\,(\hat{\mu}_1) &= \sigma^2(ad^2+1), \\ \text{var}\,(\hat{\mu}_2) &= \sigma^2(a+d^2). \end{aligned}\right\} \tag{6.2}$$

These can be used safely only if each B_i^2 is large relative to its variance, that is to say if both σ^2 and $a\sigma^2$ are small relative to unity. The weighted mean, which can be written

$$\bar{\mu} = d\left(\frac{d^2-1}{d^2+1}\right)\left(\frac{a-1}{a+1}\right), \tag{6.3}$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

can then be assigned the variance

$$\text{var}\,(\bar{\mu}) = \sigma^2 \frac{(ad^2+1)(a+d^2)}{(d^2+1)(a+1)}. \tag{6.4}$$

The expressions for $\bar{\mu}$ and var $(\bar{\mu})$ have various interesting features, the most important being that for all a, d

$$-d < \bar{\mu} < d, \tag{6.5}$$

as might be expected.

This approach certainly will give approximate results close to the truth under ideal conditions. Other approximations will not be discussed here, as they are to be regarded as superseded by the Bennett–Armitage approach and its present development. For the example

$$J(\mu) = \frac{(\mu+d)^2}{a\mu^2+1} + \frac{(\mu-d)^2}{\mu^2+a}. \tag{6.6}$$

The expression in (5.3) becomes

$$J(\mu) = \frac{a+1}{a} - 2d\mu^{-1}\left(\frac{a-1}{a}\right) - \mu^{-2}\frac{(a+1)(a^2-d^2a-a+1)}{a^2} \ldots. \tag{6.7}$$

From (6.6), (6.7) the asymptote is seen to be

$$J_\infty = \frac{a+1}{a}. \tag{6.8}$$

If $a < 1$, the approach to the asymptote will be from below at $-\infty$, from above at $+\infty$; these are reversed if $a > 1$. For $a = 1$,

$$J(\mu) = 2\left(1+\frac{d^2-1}{\mu^2+1}\right), \tag{6.9}$$

and the curve is entirely above or entirely below the asymptote as $d > 1$ or $d < 1$. At $\mu = 0$,

$$J(0) = d^2 J_\infty. \tag{6.10}$$

Attention can be restricted to $a \geqslant 1$, because an element of symmetry is present: replacement of a by $1/a$ changes $J(\mu)$ to $aJ(-\mu)$, so that the same form of dependence on μ is retained. When $d > 1$ and $a = 1$, the test statistic for heterogeneity of the two assays is

$$\chi^2_{[1]} = 2/\sigma^2$$

irrespective of the value of d, a somewhat surprising result.

For $a > 1$, in the limiting case of $d = 0$ the curve is wholly below the asymptote with a single minimum at $\mu = 0$. As d is increased, the curve develops a small 'bulge' above the asymptote at large positive μ. Only when d is moderately large do two minima appear. Fig. 2 shows the situation when $d = 4$ for various a. As is evident from Fig. 2, for $d = 4$ and $a = 1$ the function has no minimum at finite μ. The only 'point estimate' of μ is $\pm\infty$, and any interval calculated is either infinite except for exclusion of a central segment or the whole range from $-\infty$ to $+\infty$. This particular case is interesting as it is one that could easily occur in parallel line assays with two doses of each preparation. For one such assay, with appropriate conventions on scales (Finney, 1964, §5.3),

$$\text{var}\,(B_i) = \text{var}\,(D_i), \quad \text{cov}\,(B_i, D_i) = 0;$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

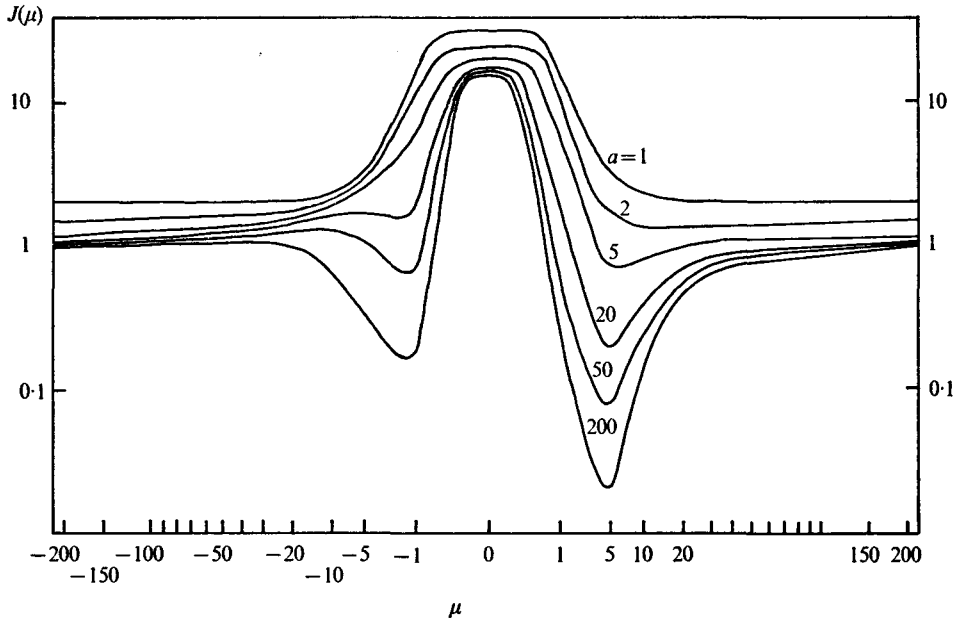

Fig. 2. The function $J(\mu)$ for $k = 2$ and

i	B_i	D_i	z_i	u_i	v_i	w_i
1	1	−4	0	a	1	0
2	1	4	0	1	a	0

The curves shown are for $a = 1, 2, 5, 20, 50, 200$.

for a pair of assays, the situation could chance to be that of equation (6.6), except for a multiplicative factor that has no effect on the form of the curve. Thus under H_{k-1} a pair of 4-point assays could give the apparently absurd conclusions described above. The position would be entirely different under a 'stronger' hypothesis that requires

$$\beta_1 = \beta_2, \quad \delta_1 = \delta_2,$$

instead of merely $\mu_1 = \mu_2$. Obviously one would then form, for the simple symmetric case,

$$\hat{B} = (B_1 + B_2)/2,$$
$$\hat{D} = (D_1 + D_2)/2,$$

and estimate μ as $\hat{D}/\hat{B}$. Since numerator and denominator are linear in the observations, an interval is determined by a standard application of Fieller's theorem, and indeed the whole problem is reduced to that of $k = 1$. The distinction between H_{k-1} and a hypothesis that states relations between the β_i and between the δ_i is clearly important.

As soon as a is increased (still with $d = 4$), the pattern changes. Thus for $a = 2$, when $\mu < 0$ the curve is similar to that for $a = 1$ except for the lower asymptote, a steady increase to a maximum high above the asymptote occurring near $\mu = 0$. As μ increases, the decline is rapid; $J(\mu)$ falls below the asymptote near $\mu = 6{\cdot}0$, reaches a minimum near $\mu = 12{\cdot}0$, and slowly increases again to approach the

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

asymptote from below. Thus, despite the separate estimates $\hat{\mu}_1 = -4{\cdot}0$, $\hat{\mu}_2 = 4{\cdot}0$, the combined estimate is approximately $\hat{\mu} = 11{\cdot}6$.

Further increase in a at first leaves the form unchanged, but makes $\hat{\mu}$ more reasonable yet still outside the interval (μ_1, μ_2); it is about 6·0 for $a = 5$, about 5·0 for $a = 20$, and about 4·5 for $a = 50$. A new phenomenon appears from about $a = 20$ onwards, the appearance of a second minimum. Thus at $a = 20$ there is a slight dip in the curve at $\mu = -1{\cdot}5$ and a well-marked minimum near $\mu = 5{\cdot}0$. At $a = 200$, the first minimum is still near $\mu = -1{\cdot}5$ but is now well below the asymptote, and the second is still near 4·5 where $J(\mu)$ comes down close to zero. By inspection of (6·6), when a is very large and μ finite, $J(\mu)$ is approximated by

$$[(1+4\mu^{-1})^2+(\mu-4)^2]/a;$$

for large enough a, this can be made to approach closely to zero at $\mu = -4{\cdot}0$ and at $\mu = 4{\cdot}0$.

The example is a little absurd for real assays. It implies that the regression coefficients are estimated as $1{\cdot}0 \pm a^{\frac{1}{2}}\sigma$ and $1{\cdot}0 \pm \sigma$: the design or replication must be rather inadequate if these coefficients have standard errors equal to or exceeding the standard deviation per response. However, the situation in which all proportionalities are the same except that standard errors are much smaller is algebraically equivalent to one in which σ^2 is taken numerically smaller. For interest, look first at $a = 2$ with $\sigma^2 = 0{\cdot}5$ (still with $d = 4$ throughout). The weighted mean formulae, (6·3) and (6·4), yield

$$\bar{\mu} = 1{\cdot}18 \pm 2{\cdot}41,$$

a reasonable enough estimate but with a standard error so large as to suggest that the estimate is useless. The *minimum minimorum* of $J(\mu)$, at $\hat{\mu} = 11{\cdot}6$ as already noted, gives

$$J(\hat{\mu})/\sigma^2 = 2{\cdot}65;$$

tested as $\chi^2_{[1]}$, this raises no doubts about H_{k-1}. The 0·95 limits to be assigned to μ are obtained by ruling a horizontal at $3{\cdot}84\sigma^2$ above $J(\hat{\mu})$, that is to say at

$$J(\mu) = 3{\cdot}24,$$

and taking the interval to consist of all μ for which $J(\mu)$ is smaller. This interval is the whole continuum except for that part from about $-4{\cdot}5$ to about $+2{\cdot}6$. Now suppose that σ^2 is much smaller, say $\sigma^2 = 0{\cdot}02$. The only change for the weighted mean is that now

$$\bar{\mu} = 1{\cdot}18 \pm 0{\cdot}48.$$

Again $\hat{\mu} = 11{\cdot}6$, but

$$J(\hat{\mu})/\sigma^2 = 66{\cdot}2.$$

This enormous value for $\chi^2_{[1]}$ must raise the gravest doubts about H_{k-1}: the data strongly conflict with the assertion that both assays are estimating the same true relative potency. If belief in H_{k-1} is too firmly entrenched to be shaken by this evidence, the 0·95 interval for μ will consist of those values for which $J(\mu)$ is less than 1·40. The interval therefore is from 7·0 to about 30·0. In both instances, the weighted mean is totally misleading.

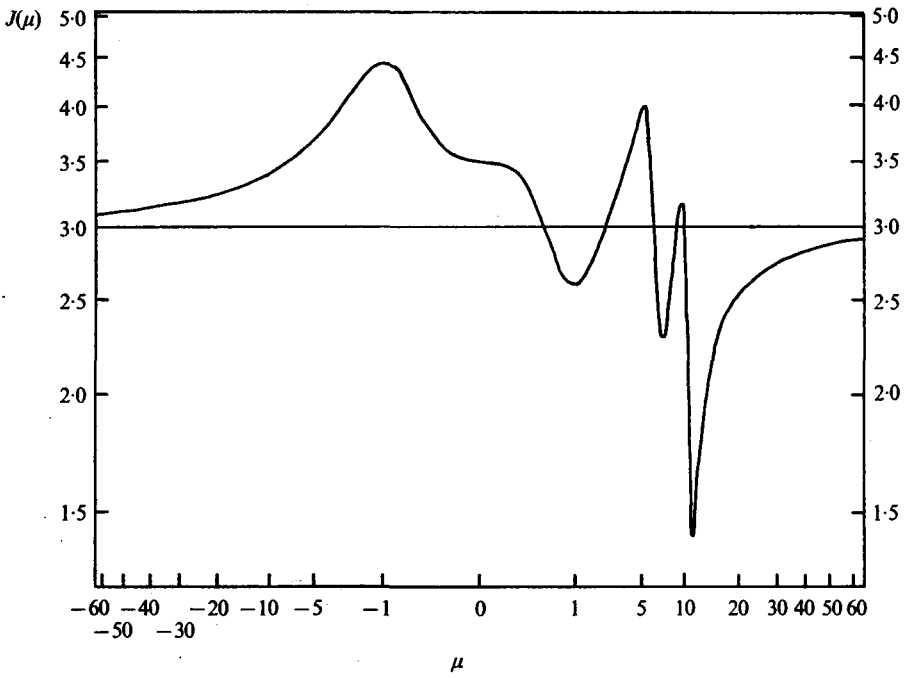

Fig. 3. The function $J(\mu)$ for $k = 3$ and

i	B_i	D_i	z_i	u_i	v_i	w_i
1	1	1	0	1	1	0
2	1	6	0	1	37	−6
3	1	11	0	1	101	−10

Enough has been said about this particular $J(\mu)$, which has little intrinsic interest. If so simple a conjunction of design parameters and data, with only two component assays, can produce such a variety of curves and point and interval estimates, a less symmetric pattern and $k > 2$ may generate far more complexities. A number of other special cases have been explored, and these have demonstrated that the various types of phenomena may occur simultaneously in the manner exemplified by Fig. 3.

Theory and numerical examples combine to show that the possibility of a probability set consisting of several discrete segments must be accepted. As stated earlier, however, real data for a problem in which H_{k-1} is a reasonable hypothesis will seldom produce apparently absurd estimates and limits at moderate probabilities.

7. A REAL EXAMPLE

Smith, Marks, Fieller & Broom (1944) present data from four assays of the same preparation of insulin. These data have been considered also by Finney (1964, §14·6; see also §10·3 for a detailed study of one of the assays) and by Bennett (1962). All assays had twin cross-over designs with three rabbits on each of the four sequences of doses. The doses of standard insulin were 1 and 2 i.u./ml., and the test doses were prepared so as to be equally potent on the provisional assumption that

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

Table 1. *Summary of four assays of the same preparation of insulin*

Assay number, i	B_i	D_i	u_i	v_i	w_i	Residual: degrees of freedom	Residual: sum of squares
1	3·500	0·833	3/64	3/16	−1/96	7	70·67
2	3·262	−3·942	1/24	1/6	0	8	151·87
3	5·584	−0·432	1/24	1/6	0	8	267·32
4	5·759	2·150	1/24	1/6	0	8	325·40

the potency of the test preparation was 22 i.u./mg. The response was the percentage reduction in blood sugar with an adjustment for initial blood sugar level. The analysis of each assay proceeds on the usual assumption of a linear relation between response and log dose. In the first assay one rabbit was lost; the appropriate method of analysis is described in §10·8 of Finney (1964).

Table 1 summarizes the essential information for the four assays. The residual mean squares do not differ significantly, and we shall use the pooled mean square

$$s^2 = 815{\cdot}26/31$$
$$= 26{\cdot}299$$

with 31 degrees of freedom, as the best estimate of residual variance. The four estimates of slope, B_i, also do not differ significantly, but we shall play safe by assuming that the true slopes, β_i, may differ. (The slopes in Table 1 are measured in terms of half the log dose interval, i.e. 0·1505 log units.) Since the nominal doses of test and standard are the same, we may take $z_i = 0$.

The function $J(\mu)$ is, from (5·1) and Table 1,

$$\begin{aligned} J(\mu) = {} & 192\,(3{\cdot}500\mu - 0{\cdot}833)^2/(9\mu^2 + 4\mu + 36) \\ & + 24(3{\cdot}262\mu + 3{\cdot}942)^2/(\mu^2 + 4) \\ & + 24(5{\cdot}584\mu + 0{\cdot}432)^2/(\mu^2 + 4) \\ & + 24(5{\cdot}759\mu - 2{\cdot}150)^2/(\mu^2 + 4). \end{aligned}$$

By direct evaluation of this function by computer, for a range of values of μ, the unique minimum is found to occur at

$$\hat{\mu} = -0{\cdot}0032.$$

This is the maximum-likelihood estimate of the common value of μ, and corresponds to an estimated relative potency (in i.u. per mg.) of

$$22{\cdot}0 \text{ antilog } (-0{\cdot}0032 \times 0{\cdot}1505) = 22{\cdot}0.$$

At this minimum, $J(\hat{\mu}) = 125{\cdot}79$. As a heterogeneity test, from (4·3), we can take

$$Z^*_{k-1} = J(\hat{\mu})/s^2 = 125{\cdot}79/26{\cdot}299$$
$$= 4{\cdot}78$$

https://doi.org/10.1017/S0022172400055030 Published online by Cambridge University Press

as a χ^2 variate with 3 degrees of freedom ($P \simeq 0{\cdot}2$) showing no significant difference between the potency estimates for the four assays. Alternatively, as discussed in the paragraph following (4·5), we could take

$$J(\hat{\mu})/3s^2 = 1{\cdot}59$$

as F with 3 and 31 degrees of freedom, giving virtually the same result.

Probability limits for μ follow from (5·6), or by replacing $\chi^2_{[1]}$ by $F(1, 31)$ as indicated below (5·6). The 5% tabular value of $F(1, 31)$ is 4·16, and 0·95 limits for μ are thus determined by

$$\begin{aligned} J(\mu) &= 125{\cdot}79 + (4{\cdot}16)(26{\cdot}299) \\ &= 235{\cdot}18. \end{aligned}$$

Tabulation of $J(\mu)$ shows that this value corresponds to $\mu = -0{\cdot}489$ and $0{\cdot}486$, with limits for the potency at 18·6 and 26·0 i.u. per mg. Finney (1964, §14·6), using a different method of combining assay results, obtained a similar estimate, 21·4, with limits of 18·0 and 25·3 i.u./mg. The function $J(\mu)$ is well behaved in this example, and the possible complications discussed in the previous section do not arise.

We are indebted to Mr D. A. Williams for drawing our attention to various references relating to linear functional relations.

REFERENCES

ARMITAGE, P. (1970). The combination of assay results. *Biometrika*, **57**, 665–6.

ARMITAGE, P. & BENNETT, B. M. (1974). Maximum likelihood solutions for the combination of relative potencies. *Journal of Hygiene* **73**, 97–9.

ARMITAGE, P., BAILEY, J. M., PETRIE, A., ANNABLE, L. & STACK-DUNNE, M. P. (1974). Studies in the combination of bio-assay results. *Biometrics* **30**, 1–10.

BARNETT, V. D. (1966). Evaluation of maximum likelihood estimator where the maximum likelihood equation has multiple roots. *Biometrika* **53**, 151–65.

BEALE, E. M. L. (1960). Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society* B **22**, 41–88.

BEALE, E. M. L. (1966). Contribution to discussion on Sprent (1966).

BENNETT, B. M. (1962). On combining estimates of relative potency in bioassay. *Journal of Hygiene* **60**, 379–85.

BENNETT, B. M. (1963*a*). On combining estimates of a ratio of means. *Journal of the Royal Statistical Society* B **25**, 201–5.

BENNETT, B. M. (1963*b*). Slope ratio assays and confidence limits. *Metrika* **7**, 117–20.

BROWN, R. L. (1957). Bivariate structural relation, *Biometrika* **44**, 84–96.

DOLBY, G. R. (1972). Generalized least squares and maximum likelihood estimation of non-linear functional relationships. *Journal of the Royal Statistical Society* B **34**, 393–400.

EDWARDS, A. W. F. (1972). *Likelihood*. Cambridge University Press.

FINNEY, D. J. (1964). *Statistical Method in Biological Assay*, 2nd ed. London: Griffin.

SMITH, K. W., MARKS, H. P., FIELLER, E. C. & BROOM, W. A. (1944). An extended cross-over design and its use in insulin assay. *Quarterly Journal of Pharmacy and Pharmacology* **17**, 108–17.

SPRENT, P. (1966). A generalized least-squares approach to linear functional relationships. *Journal of the Royal Statistical Society* B **28**, 278–97.

SPRENT, P. (1969). *Models in Regression and Related Topics*. London: Methuen.

VILLEGAS, C. (1964). Confidence regions for a linear relation. *Annals of Mathematical Statistics* **35**, 780–8.