Ann. Rev. Pharmacol. Toxicol. 1987. 27:385–97

STATISTICAL METHODS AND APPLICATIONS OF BIOASSAY

Eugene M. Laska

Nathan S. Kline Institute for Psychiatric Research, Statistical Sciences and Epidemiology Division, Orangeburg, New York 10962

Morris J. Meisner

New York University Medical Center, Department of Psychiatry, New York, New York 10016

INTRODUCTION

A bioassay experiment is designed to estimate the relative potency of a test compound (T) to a standard compound (S). The relative potency of T to S is defined as the dose of T that produces the same biological response as does a unit dose of S (1).

Traditionally, relative potency has been estimated for a test compound that is a dilution of the standard. Currently, two different compounds that have parallel dose-response curves over some range of doses are rountinely assayed. This case is known as a comparative assay.

A bioassay experiment may be either quantal or quantitative, direct or indirect. If the response measure is binary, the assay is said to be quantal (2). Otherwise, it is quantitative. In a direct assay the threshold dose required for response is determined for each experimental unit. Thus, the observed data are dose units. In an indirect assay the experimental unit receives one or more specified doses of the preparation and the observed data may be either quantal or quantitative responses. Depending on the experimental design, several dose levels of T and S are given to the same or different experimental units. The former experiment is called a *crossover,* and the latter, a *parallel group* or *completely randomized design.*

In this chapter we discuss only indirect quantitative assays. We first review the elementary theory developed over the past forty years for the univariate

0362-1642/87/0415-0385$02.00

case. Next, we summarize recent work that extends the theory to the multivariate situation. This theory enables the computation of a single estimate of relative potency based on many measures of outcome. It also provides a method for pooling results across experiments. We then apply the theory to the problem of determining whether two preparations are bioequivalent and whether mixtures of compounds are synergistic, additive, or antagonistic.

UNIVARIATE BIOASSAY

The statistical analysis of a bioassay experiment requires a model that relates the average response to the dose of the preparations. If the average response is linearly related to the dose, and the line passes through the origin, the model is called a *slope-ratio assay*. However, a more commonly used model is the *parallel line assay*. In this assay response is linearly related to the log dose, and the line need not pass through the origin. The mathematical setup of the parallel line assay is as follows. Let the dose of the standard (test abbreviations in parentheses) preparation be denoted by z_S (z_T) and the log dose by $x_S = \log z_S$ ($x_T = \log z_T$). Let the response variable be denoted by Y_S for S and Y_T for T. The statistical model assumes that

$$Y_S = \alpha_S + \beta x_S + \epsilon_S \quad (Y_T = \alpha_T + \beta x_T + \epsilon_T), \qquad 1.$$

where ϵ_S (ϵ_T) is distributed normally with mean 0 and variance σ^2 for each value of x_S (x_T). In other words, $\alpha_S + \beta x_S$ is the average response due to the dose x_S of S, and $\alpha_T + \beta x_T$ is the average response due to the dose x_T of T. If T is a dilution of S, a proportion, ρz_T, of the test preparation will have an effect equal on the average to a dose, z_S of the standard preparation for every value of z_T in the range of doses for which the model holds. The relationships

$$z_S = \rho z_T$$

and

$$\alpha_S + \beta x_S = \alpha_T + \beta x_T$$

define the constant ρ, the relative potency of T to S. It follows that

$$\alpha_S + \beta \log \rho z_T = \alpha_T + \beta \log z_T$$

or

$$\log \rho = (\alpha_T - \alpha_S)/\beta. \qquad 2.$$

This expression defines the log of the relative potency in terms of the parameters α_T, α_S, and β of the *Parallel Line Model* 1. The geometric interpretation of the quantity, log ρ, is given in Figure 1.

To compute the point estimator of logρ, the point estimators, $\hat{\alpha}_S$, $\hat{\alpha}_T$, and $\hat{\beta}$ of the quantities α_S, α_T, and β are necessary. These numbers are obtained by using ordinary least squares regression and then substituting the results into *Equation 2*. These regression estimators are computed in a manner similar to the computation of slopes and intercepts in a straight-line regression problem, but in this case, because of parallelism, the slopes of the two lines are constrained to be equal.

Also, a confidence interval for logρ is useful. (A statistical procedure is said to produce a 95% confidence interval for a parameter when, if repeated many times, the true value of the parameter would be included 95% of the time. The values defining the interval depend on the observed responses in each repetition of the experiment.)

Fieller's Theorem (3, 4) provides a method used to obtain a confidence interval for a ratio of parameters when the estimators of both the numerator and the denominator are normally distributed random variables. Its application to bioassay derives from *Equation 2*, which expresses log relative potency as a ratio of the parameters $\alpha_T - \alpha_S$ to β from *Model 1*.

Let $\hat{\delta} = \hat{\alpha}_T - \hat{\alpha}_S$. From elementary regression theory it follows that $\hat{\delta}$ is a normal random variable with mean $\alpha_T - \alpha_S$, and $\hat{\beta}$ is a normal random variable with mean β. Then $E(\hat{\delta} - \mu\hat{\beta}) = 0$, where the notation E indicates the expected value of the random variable. In such a situation, Fieller's Theorem asserts that

$$(\hat{\delta} - \mu\hat{\beta})^2 \leqslant F_{1,\nu}(k_1\mu^2 - 2k_2\alpha 3 + k_3)s^2$$

defines a confidence region for $\mu = \log\rho$. Here s^2 is the point estimator of σ^2 obtained from *Model 1;* k_1, k_2, and k_3 are constants that depend on the log doses and the number of test and standard observations; $\nu = N-3$, where N is the total number of observations, and $F_{1,\nu}$ is the 95th percentile of an F distribution with 1 and ν degrees of freedom. The confidence interval is determined by the roots, A and B, of the quadratic equation

$$a\mu^2 + b\mu + c = 0,$$

where $a = \hat{\beta}^2 - F_{1,\nu}k_1$, $b = -2\hat{\delta}\hat{\beta} + 2k_2F_{1,\nu}$ and $c = \hat{\delta}^2 - F_{1,\nu}k_3$. If $a > 0$ then the confidence region is the interval [B,A]. But if $a<0$ and $b^2-4ac > 0$, then the confidence region is the exterior of the interval [B,A], i.e. the set $\log\rho \leqslant B$ and $\log\rho \geqslant A$. A third possibility is that the confidence region is the entire straight line. This uninformative consequence of Fieller's Theorem occurs when $a < 0$ and $b^2 - 4ac < 0$. No other cases are possible (5).

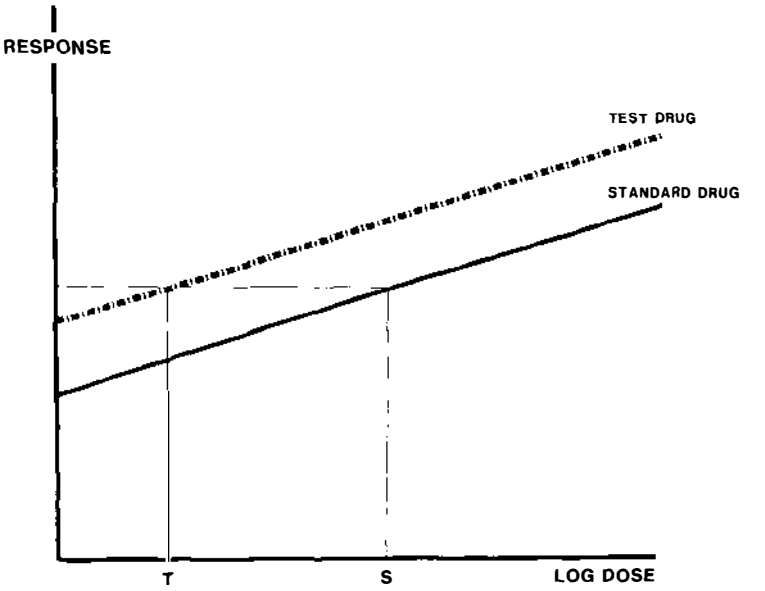

Figure 1 The length of the interval (T, S) is the log relative potency. The dose T of the test drug produces the same response as the dose S of the standard drug.

It is standard procedure to test the validity of the model of *Equation 1* via an analysis of variance (ANOVA) (1). In Table 1 the entry "Regression" tests the null hypothesis that the slope, β, is zero. This test is important for if β is zero, the concept of relative potency has no meaning because from *Equation 1* the relative potency parameter is undefined. "Parallelism" (Table 1) tests the assumption that the dose response lines for test and standard are indeed parallel. The model of *Equation 1* assumes that the average response of Y_S depends linearly on the log dose, x_S. The true relation between Y_S and x_S may be more complicated. They could be related by higher order terms, such as a quadratic involving the square of x_S. Also in Table 1, "Linearity" tests the possibility that a linear model is an inadequate representation of the dose-response relationship. The entry "Preparations" examines whether the same range of effectiveness is studied in the two treatments. It tests the null hypothesis that the weighted-population mean response of all doses of the standard equals that of all doses of the test used in the experiment. If so, the assay may still be valid. The magnitude of the difference indicates how great an extrapolation in the effect range is involved.

Although the ANOVA is traditionally used to test validity, it is somewhat inconsistent with the analogous tests based on the model of *Equation 1*. Using likelihood-ratio theory all tests can be based on *Equation* 1, thereby avoiding any possibility of inconsistencies. As an illustration, the estimator of σ^2

Table 1 Analysis of variance for parallel line assay based on a completely randomized (parallel groups) design

Source of variation	Degrees of freedom[a]
Between doses	$K_S + K_T - 1$
Regression	1
Parallelism	1
Preparations	1
Linearity	$K_S + K_T - 4$
Within doses	$N_S + N_T - K_S - K_T$
Total	$N_S + N_T - 1$

[a] N_S (N_T) is the number of experimental units receiving the standard (test) preparation, and K_S (K_T) is the number of different doses of the standard (test) preparation.

(required for instance by Fieller's Theorem) resulting from use of the ANOVA, is not the same as that resulting from the use of *Equation 1*. The ANOVA method ignores, whereas the likelihood-ratio method incorporates, the dose-response relationship (6).

COMBINATION OF UNIVARIATE ASSAYS

The problem of combining bioassays (1, 7–9) has received renewed attention in recent years. Suppose that several independent bioassay experiments designed to estimate the relative potency of the same test and standard treatments are performed. Each one produces a point and 95% confidence interval estimator of ρ. How could a single point and interval estimate of ρ be produced on the basis of all of the data? Clearly, such combined estimates should be more accurate than those produced by the individual experiments. By application of the method of maximum-likelihood estimation, a combined estimator can be obtained (7, 9). Two features of the combined estimator may be mentioned: (*a*) the combined estimator of ρ depends not only on the individual estimates of ρ but on the sample sizes, experimental design, estimated residual error, and the statistics $\hat{\alpha}_T$, $\hat{\alpha}_S$, and $\hat{\beta}$ obtained in each of the experiments; and (*b*) the estimator of log ρ is obtained from the solution to a polynomial equation whose degree is greater than two. This latter situation does not occur in the problems mentioned previously. For example, $\hat{\alpha}_T$ is obtained from a linear equation. The Fieller confidence interval is obtained by solving a quadratic equation. The new mathematical complication occurs because estimation of the common relative potency is a nonlinear problem. Such problems in mathematical statistics generally require the solution of complicated nonlinear and nonquadratic equations.

Combining many bioassays involves first testing the validity assumptions in the individual studies. Even if a few such studies fail, they need not necessarily be excluded from the analysis. Complications arise when the slope of the regression lines of one or two of the studies has a sign that is opposite to those of the other studies. The additional hypothesis requiring investigation is whether the drugs share a common relative potency across experiments. This hypothesis can also be tested by the likelihood-ratio approach. Mathematical complications also arise here because such a hypothesis is not linear. In particular, instead of minimizing a sum of squares (as in straight-line regression theory), finding the likelihood-ratio test in this situation requires minimizing a function of the form

$$J(\mu) - \sum_{i-1}^{I} (\hat{\delta}_i - \mu\hat{\beta}_i)^2/(p_i\mu^2 + q_i\mu + r_i),$$

where p_i, q_i, and r_i are constants depending on the log doses in the i^{th} experiment. I is the number of investigators and $\mu = \log\rho$. The value of μ that minimizes $J(\mu)$ is $\hat{\mu}$, the combined maximum-likelihood estimator of log relative potency. The solution can be found numerically. Various values of μ are systematically substituted into the above equation until the one that produces the minimum value of $J(\mu)$ is found. Another approach to finding the minimum is possible because $J(\mu)$ can be conveniently expressed as a ratio of two polynomials: $J(\mu)=P_1(\mu)/P_2(\mu)$. Standard techniques of differential calculus can then be used to obtain $\hat{\mu}$ as the root of a polynomial equation.

Finally, using likelihood-ratio methods, a Fieller-like confidence region (C) for $\log\rho$ may be obtained. In general, likelihood-ratio theory requires the specification of two hypotheses, H_0 and H_1. In the above case, H_0 is the hypothesis that, across all experiments, the log relative potency is equal to a specified value, $\mu_0 = \log\rho_0$. C is the acceptance region of H_0, i.e. the set of all μ_0 for which H_0 is not rejected. An exact method and an asymptotic method, both of which use the same H_0, have been developed. They differ in the choice of H_1, the "alternative hypothesis." If H_1 is the hypothesis that all experiments share an unknown common relative potency, then asymptotic or large-sample distribution theory must be used. If H_1 is the general linear model of *Equation 1*, a common relative potency across all experiments is not assumed, which results in the applicability of exact or small-sample distribution theory. The two approaches lead to different denominators in the likelihood ratio used to compute confidence regions. Nevertheless, the mathematical forms of the two confidence regions are very similar. Note that the word *region* rather than *interval* is used, for C may consist of one interval or many intervals. The possibility of many finite intervals in the confidence region

does not occur when a confidence region for $\log\rho$ is computed for one bioassay. When the data is not pathological, however, C probably consists of one interval. If C is the asymptotic confidence region, then $\hat{\mu}$, the combined estimator of $\log\rho$, always belongs to C. If C is is the exact confidence region, it may be empty. This situation, too, does not occur with one bioassay. However, when the exact confidence region C is not empty, $\hat{\mu}$ always belongs to it.

SINGLE AND COMBINED MULTIVARIATE BIOASSAYS

Several papers have discussed multivariate bioassay (6, 10–13). Multivariate bioassay estimates relative potency when the observation of an experimental unit is not a single value but a vector of multiple responses representing different, but possibly correlated, outcomes. As in a univariate assay, each of the responses of the two drugs S and T are assumed to follow a linear model. For the i^{th} response variable, we assume the model

$$Y_{S,i} = \alpha_{S,i} + \beta_i \log z_S + \epsilon_{S,i} \quad (Y_{T,i} = \alpha_{T,i} + \beta_i \log z_T + \epsilon_{T,i}), \qquad 3.$$

where, as before, z_T and z_S are the log doses. For the i^{th} response variable, *Model 3* is identical to *Model 1,* the model in the univariate situation. Here, however, if there are, for example, three responses, the covariances of the error terms $\epsilon_{T,1}$, $\epsilon_{T,2}$, and $\epsilon_{T,3}$ play a major role. The vector $\epsilon_T = (\epsilon_{T,1}, \epsilon_{T,2}, \epsilon_{T,3})$ is assumed to be multivariately normally distributed with mean vector (0, 0, 0) and unknown variance-covariance matrix Σ. Note that, as in the univariate case, the standard and test formulations for each fixed i are assumed to have the same slope, β_i, although these slopes may be different for different response measures. If T is a dilution or new formulation of S, the relative potency should be the same, irrespective of the response measure used. The main goals of a multivariate bioassay are to (*a*) test the validity of the model of common slope; (*b*) test the hypothesis that there is a common relative potency for each variable; (*c*) estimate the scalar quantity, the common relative potency; and (*d*) determine a confidence interval for the common relative potency.

If two or more independent bioassay experiments are performed on T and S, as in the univariate case, each experiment separately yields a relative potency. In addition to the above points, another goal of combining multivariate bioassays (13–15) is (*e*) to analyze the data of all the experiments simultaneously to obtain a single point and interval estimator of relative potency.

The mathematics of these problems are more difficult than those of corresponding problems that arise in the univariate situation. Goal (*a*) is easiest to

reach because it can be formulated as a linear hypothesis with respect to a multivariate model, at the same time being somewhat more general than the multivariate *Model 3*. The theory of tests of such linear hypotheses is well known. Thus, a test of (*a*) uses the F-test statistic. In short, (*a*) is treated as a special case of standard linear multivariate statistical theory. Hypothesis (*b*), on the other hand, does not arise univariately and is a nonstandard hypothesis testing situation because it is a nonlinear hypothesis. Asymptotic, i.e. large-sample, mathematical techniques are unavoidable. In the multivariate bioassay case, likelihood-ratio theory facilitates a large-sample test of the hypothesis of common relative potency. The same technique yields a maximum-likelihood estimator of the common relative potency and a corresponding 95% confidence interval. In a single multivariate bioassay experiment the methods that generate the tests and estimators described in (*a*)–(*d*) surprisingly turn out to require nothing more complicated than solving quadratic equations. However, analysis (*e*), combining several multivariate bioassays, involves solving higher order polynomial equations, as in the situation that arises from combining several univariate bioassays. A maximum-likelihood solution to the problem of combining multivariate bioassays has just recently appeared. The mathematical detail of the statistical analyses of these experiments is given in the references (13, 15) cited above. In the following sections we present some applications of these methods.

BIOEQUIVALENCE

There is considerable current interest in the interchangeability of generic and brand-name compounds (16–28). Suppose there is a standard drug S and a test drug T, which is a new formulation of the standard. The statistical problem is how to determine whether the two formulations are equivalent. The experimental procedure is accomplished in vivo, using a crossover or a completely randomized design. After either a single dose or multiple doses are administered to a pseudosteady state, blood samples are obtained to find the serum concentration over a single dose interval. All inferences about bioequivalence are made on the basis of this single dose.

The concentration curves are usually summarized by three measures: (*a*) the area under the curve *(AUC);* (*b*) the peak concentration (C_{max}); and (*c*) the time to achieve the peak concentration (T_{max}). Usually each of the three measures (or a transformation such as their logarithms) is assumed to follow a normal distribution. The statistical hypothesis of interest is whether, for example, the expected values of the *AUC* given the standard drug S, *E[AUC|S]*, and the analogous quantity, *E[AUC|T]*, are equal. In practice, equality is not necessary, and if each of the three measures are similar, the formulations are said to be bioequivalent. Note that in the usual statistical

situation the hoped for result is rejection of the null hypothesis, whereas in this case bioequivalence is concluded when the null hypothesis is not rejected. One must be sure, however, that the statistical power is adequate to reject if the drugs are not bioequivalent. Procedures for judging bioequivalence include classical statistical hypothesis testing, confidence intervals for the difference of, $E[AUC|S]-E[AUC|T]$, confidence intervals for the ratio $E[AUC|T]/E[AUC|S]$, and Bayesian methodology. Typically, the above methods are carried out univariately, that is, separately for each of the three measures.

One widely used rule accepts the proposition that two formulations are bioequivalent if the serum summary parameters of the test formulation are within 20% of those of the standard formulation. Let U_T and U_S denote the summary serum concentration parameters, and let $\Theta = U_T/U_S$. If with probability 0.95, $0.8 \leq \Theta \leq 1.2$, then the formulations are said to be bioequivalent under this rule.

Several different statistical approaches to obtaining confidence intervals have been suggested. Fieller's Theorem could be applied, and if the resulting interval falls within 0.8 to 1.2, then bioequivalence is accepted. Other authors have recommended methods for forming a confidence interval for Θ that is symmetric about 1. These approaches use the fact that there are many pairs of scalars for the t distribution with ν degrees of freedom that may be used to form a 95% confidence interval. In contrast to the usual scalars that are determined a priori, e.g. ± 1.96, these approaches choose limits that are data dependent to obtain a symmetric confidence interval.

Various post hoc statistical procedures are also used to perform power calculation to estimate the chance of detecting a difference, as a function of the true difference. These methods are post hoc in that the estimate of variability obtained from the results of the experiment is used in the power calculation.

The three measures usually used to summarize the serum concentration curves are obviously not three independent random variables but are rather components of a multivariate-response vector. It is clearly inappropriate to treat the responses univariately, since such a procedure results in several different measures of bioequivalence and ignores the covariance relationships between the components of thc response vector. Therefore the appropriate approach is multivariate, in which a single measure of bioequivalence and a 95% confidence interval is obtained. In the case of measures that are ratios, a multivariate analogue of Fieller's Theorem can be applied. Also, rather than use the three summary measures of the concentration time curve, it may be more appropriate to use the actual observations of serum level at the times that they are obtained.

If the two formulations are found to be bioequivalent at one dose, the

question of bioequivalence at a different dose is generally not tested in vivo. However, if there is therapeutic interest in a range of doses, it is logical to study the bioavailability over the full range. If the experimental design includes multiple doses of S and T, then the application of bioassay techniques is immediate and the scalar quantity relative potency is the natural indicator of bioequivalence. As mentioned above, if the confidence region for relative potency lies between 0.8 and 1.2, the two formulations may be said to be bioequivalent.

Finally, when there are several experiments it is important to have a single reply to the question of bioequivalence. Here the method of combining multivariate bioassays from two or more experiments provides a setting in which a single common relative potency and a corresponding 95% confidence region provide an unambiguous measure of bioequivalence.

MIXTURES

When two drugs, denoted by D_1 and D_2, are mixed in known quantities to produce a combination agent, its effectiveness is frequently compared to the effects of the constituent drugs acting separately. Suppose that the drugs are equipotent. If, for all doses x_1 and x_2, the effect of the mixture of dose x_1 of D_1 and dose x_2 of D_2 equals the effect of D_1 (or D_2) at dose x_1+x_2, the mixture of the two drugs is said to be *additive*. Also, the two drugs are said to be *synergistic* or *antagonistic* according to whether the effect of the mixture is greater than or less than the effect of D_1 (or D_2) at dose x_1+x_2. A curve of the doses (x_1, x_2) for which the effect of the mixture is constant is an *isobole,* and a collection of these curves, at different effect levels, is an *isobologram* (29). The type of joint action described by the additive model is *simple similar action.*

More generally, suppose the relative potency of D_1 to D_2 is $\mu = \log \rho$. Further, assume that the effect of the mixture is a random variable that is normally distributed. The null hypothesis H_0 states that the mixture is additive, i.e. the effect random variable has a mean equal to $\alpha + \beta\log(x_1 + \rho x_2)$. Notice that the model is statistical and makes no attempt to represent the biological action of the drugs (30–33). If the mixture is additive, point and interval estimators for the parameters may be obtained using maximum-likelihood theory, although numerical methods and asymptotic theory are required (34, 35).

The question of additivity versus synergism or antagonism may be tested as a statistical hypothesis using likelihood-ratio theory. Let the j^{th} dose of the i^{th} mixture be (x_{1ij}, x_{2ij}) where the i^{th} mixture has the property that $x_{2ij} = \gamma_i x_{1ij}$. A common design is to have n_{ij} subjects receive the i^{th} mixture at the j^{th} dose level.

To perform the likelihood-ratio test, an alternative hypothesis, H_1, needs to be specified. One natural alternative is a model in which for each dose mixture of ratio γ_i, there is a separate constant $\mu_i = \log\rho_i$, such that the effects of the mixture of dose x_{1ij} of D_1 and dose x_{2ij} of D_2 is equivalent to the effect of dose $x_{1ij} + \mu_i x_{2ij}$ of D_1. For the i^{th} mixture, the expected response is

$$\alpha + \beta\log(x_{1ij} + \mu_i x_{2ij}).$$

Substituting $x_{2ij} = \gamma_i x_{1ij}$, the expected response is equal to

$$c_i + \beta\log(x_{1ij} + x_{2ij}),$$

where

$$c_i = \alpha + \beta\log[(1 + \mu_i\gamma_i)/(1 + \gamma_i)].$$

The first step is checking that the dose-response lines of the mixture are parallel. This may be accomplished by fitting the model

$$E(\text{response}) = c_i + \beta_i\log(x_{1ij} + x_{2ij})$$

and testing the null hypothesis that the β_i are equal. If parallelism is rejected, the model of additivity is not possible and must be rejected. If parallelism is not rejected, the common slope β and c_i are estimated by least squares, and the quantities μ_i are determined from them in the same fashion as in the calculation in *Equation 2*.

Another alternative, H_1, assumes no structure other than that the responses are normally distributed. In either case the test of additivity proceeds first by independently maximizing the numerator likelihood under H_0 and the denominator likelihood under H_1 and computing their ratio Λ. To check significance $-2\log\Lambda$ is compared to the 95th percentile of a chi-square distribution.

The maximum-likelihood estimator, $\hat{\mu}$, is a by-product of the above calculation. In the computation of the numerator under the additivity model H_0, the value of μ for which the maximum of the likelihood-ratio statistic is achieved is the maximum-likelihood estimator.

To find a confidence interval first test that the relative potency of D_1 to D_2 is equal to a known constant, ρ_0. Under this H_0, the least-squares estimators of the remaining parameters α and β can then be obtained by standard regression theory because the resulting model is linear. The estimators depend, of course, on the particular value of ρ_0. Using either specification of H_1, the null hypothesis of additivity can be tested using the likelihood ratio. The resulting test statistic follows an F distribution, and rejection occurs when

the statistic exceeds the 95th percentile. The 95% confidence interval then consists of all points ρ_0 for which this null hypothesis is not rejected.

ACKNOWLEDGMENTS

The authors would like to thank H. B. Kushner for his helpful comments and discussions.

Literature Cited

1. Finney, D. J. 1978. *Statistical Method in Biological Assay*. New York: MacMillan
2. Finney, D. J. 1947. *Probit Analysis. A Statistical Treatment of the Sigmoid Response Curve*. Cambridge: Cambridge Univ. Press
3. Fieller, E. C. 1940. The biological standardization of insulin. *J. R. Stat. Soc.* 137:1–53
4. Fieller, E. C. 1940. A fundamental formula in the statistics of biological assay, and some applications. 17:117–23
5. Buonaccorsi, J. 1979. Letter to the editor. *Am. Stat.* 33:162
6. Laska, E., Kushner, H. B., Meisner, M. 1985. Reader reaction: Multivariate bioassay. *Biometrics* 41:547–54
7. Bennett, B. M. 1962. On combining estimates of relative potency in bioassay. *J. Hyg.* 60:379–85
8. Armitage, P., Bennett, B. M., Finney, D. J. 1976. Point and interval estimation in the combination of bioassay results. *J. Hyg.* 76:147–62
9. Williams, D. A. 1978. An exact confidence region for a relative potency estimated from combined bioassays. *Biometrics* 34:659–61
10. Rao, C. R. 1954. Estimation of relative potency from multiple response data. *Biometrics* 10:208–20
11. Volund, A. 1980. Multivariate bioassay. *Biometrics* 36:225–36
12. Carter, E. M., Hubert, J. J. 1985. Analysis of parallel line assays with multivariate responses. *Biometrics* 41:703–10
13. Srivastava, M. S. 1986. Multivariate bioassay, combination of bioassays and Fieller's theorem. *Biometrics* 42:131–41
14. Volund, A. 1982. Combination of multivariate bioassay results. *Biometrics* 36:225–36
15. Meisner, M., Kushner, H. B., Laska, E. M. 1986. Combining multivariate bioassays. *Biometrics* 42:421–27
16. Westlake, W. J. 1972. Use of confidence intervals in analysis of comparative bioavailability trials. *J. Pharm. Sci.* 61:1340–41
17. Metzler, C. M. 1974. Bioavailability—a problem in equivalence. *Biometrics* 30:309–17
18. Westlake, W. J. 1976. Symmetric confidence intervals for bioequivalence trials. *Biometrics* 32:741–44
19. Westlake, W. J. 1979. Statistical aspects of comparative bioavailability trials. *Biometrics* 35:273–80
20. Rodda, B. E., Davis, R. L. 1980. Determining the probability of an important difference in bioavailability. *Clin. Pharmacol. Ther.* 28:247–52
21. Mandallaz, D., Mau, J. 1981. Comparison of different methods for decision-making in bioequivalence assessment. *Biometrics* 37:213–22
22. Selwyn, M. R., Dempster, A. P., Hall, N. R. 1981. A Bayesian approach to bioequivalence for the 2 × 2 changeover design. *Biometrics* 37:11–21
23. Fluehler, H., Grieve, A. P., Mandallaz, D., Mau, J., Moser, H. A. 1983. Bayesian approach to bioequivalence assessment: An example. *J. Pharm. Sci.* 72:1178–81
24. Metzler, C. M., Huang, D. C. 1983. Statistical methods for bioavailability. *Clin. Res. Pract. Drug. Reg. Affairs*, 1:109–32
25. Hauck, W. W., Anderson, S. 1984. A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J. Pharmacokinet. Biopharm.* 12:83–91
26. Rocke, D. M. 1984. On testing for bioequivalence. *Biometrics* 40:225–30
27. Rocke, D. M. 1985. Correspondence. *Biometrics* 41:561–63
28. Selwyn, M. R., Hall, N. R. 1984. On Bayesian methods for bioequivalence. *Biometrics* 40:1103–8
29. Loewe, S. 1957. Antagonisms and antagonists. *Pharmacol. Rev.* 9:237–42
30. Ashford, J. R. 1958. Quantal responses to mixtures of poisons under conditions of similar action—the analysis of uncontrolled data. *Biometrika* 45:74–88
31. Plackett, R. L., Hewlett, P. S. 1963. A unified theory for quantal responses to

mixture of drugs: the fitting of data of certain models for two non-interactive drugs with complete positive correlation of tolerances. *Biometrics* 19:517–31
32. Ashford, J. R., Smith, C. S. 1964. General models for quantal response to the joint action of a mixture of drugs. *Biometrika* 51:413–28
33. Plackett, R. L., Hewlett, P. S. 1967. A comparison of two approaches to the construction of models for quantal responses to mixtures drugs. *Biometrics* 23:27–44
34. Scaf, A. H. 1974. Statistical analysis of isoboles. *Arch. Int. Pharmacodyn.* 208:138–65
35. Darby, S. C., Ellis, M. J. 1976. A test for synergism between two drugs. *Appl. Stat.* 25:296–99