Status: Currently Official on 12-Feb-2025
Official Date: Official as of 01-Aug-2020
Document Type: General Chapter
DocId: GUID-E40D050D-B163-43F4-8D82-891AF3907355_3_en-US
DOI: https://doi.org/10.31003/USPNF_M98860_03_01
DOI Ref: h2qq0

⟨111⟩ DESIGN AND ANALYSIS OF BIOLOGICAL ASSAYS

INTRODUCTION

The potency of several Pharmacopeial articles must be determined by bioassays. The aim of this chapter is to present a concise account of certain essential biometrical procedures for bioassays in chapters or monographs of *USP–NF*, namely outlier identification, confidence intervals for relative potency measurements, and combination of independent assays. For bioassays not in *USP–NF*, other methods may be appropriate. See general information chapter *Analysis of Biological Assays* ⟨1034⟩ which may be a helpful, but not mandatory, guidance.

REJECTION OF OUTLYING OR ABERRANT OBSERVATIONS

A response that is questionable because of failure to comply with the procedure during the course of an assay is rejected. Other aberrant values may be discovered only after the responses have been tabulated, but can then be traced to assay irregularities that justify their omission. The arbitrary rejection or retention of an apparently aberrant response can be a serious source of bias. In general, the rejection of observations solely on the basis of their relative magnitudes, without investigation as to cause, is a procedure to be used sparingly. Should it be understood, either following an investigation into cause or based on practical assay experience, that an observation's discordance is unlikely to arise from a reasonable expectation of response to assay treatments, then a suspected aberrant response or outlier may be tested against one of two criteria, both of which assume that the data have an approximately normal distribution (which may be satisfied only after a suitable transformation of the original responses). Alternative statistically sound approaches to outlier detection may be used. The conditions under which outlier testing will be conducted and the criterion to be used should be specified a priori in the lab's procedures if not specified in the monograph or chapter.

Criterion 1 (Dixon's Test)

The first criterion is based on the variation within a single group of supposedly equivalent responses, such as a group of animals given a common concentration of a sample. At a confidence level of 99%, a valid observation will be rejected once in 100 trials (when the suspected outlier can occur at only one end) or once in 50 trials (when the suspected outlier can occur at either end), provided that relatively few, if any, responses within the group are identical. Arrange the responses in order of magnitude from y_1 to y_N, where N is the number of observations in the group. Compute the relative gap by using the formulas in *Table 1* below.

Table 1

Sample Size (N)	Candidate Outlier is Smallest (y_1)	Candidate Outlier is Largest (y_N)
3–7	$G_1 = (y_2 - y_1)/(y_N - y_1)$	$G_1 = (y_N - y_{N-1})/(y_N - y_1)$
8–10	$G_2 = (y_2 - y_1)/(y_{N-1} - y_1)$	$G_2 = (y_N - y_{N-1})/(y_N - y_2)$
11–13	$G_3 = (y_3 - y_1)/(y_{N-1} - y_1)$	$G_3 = (y_N - y_{N-2})/(y_N - y_2)$

If G_1, G_2, or G_3, as appropriate, exceeds the critical value in *Table 2*, for the observed N, there is a statistical basis for identifying the discordant measurement as an outlier and considering its removal. For N larger than 13, use *Criterion 2*.

In samples from a normal population, at a confidence level of 99%, gaps equal to or larger than the following values of G_1, G_2, and G_3 occur with a probability $P = 0.01$, when outlier measurements can occur only at one end; or with $P = 0.02$, when they may occur at either end.

Table 2. Test for Outlier Measurements

N	3	4	5	6	7
G_1	0.988	0.889	0.780	0.698	0.637
N	8	9	10	—	—

G_2	0.683	0.635	0.597	–	–
N	11	12	13	–	–
G_3	0.679	0.642	0.615	–	–

Criterion 2 (Grubbs, Extreme Studentized Deviate Test)

The second criterion may be used to examine for outlying values in groups of supposedly equivalent responses and may also be used in examining the set of residuals from a fitted model (linear or nonlinear) where there is constant variance. The final model (which yields the residuals for outlier detection) should include all important design variables. (For further discussion of design variables, see general information chapter *Design and Development of Biological Assays* ⟨1032⟩, which may be a helpful, but not mandatory, resource.) (Note that for application to residuals, the following is an approximation. If the statistical software provides studentized residuals, those values should be used instead of those from the following equation.) For the value, R, that is furthest from the sample mean, compute the standardized deviation Z:

$$Z = (R - \overline{R})/S$$

where $\overline{R}$ and S are the mean and standard deviation, respectively, of the set of values. For residuals from a least squares fit, such as for a parallel line assay, $\overline{R} = 0$, and S is the square root of the residual mean square from the analysis. If $|Z|$ is greater than C as determined below, then the value R is identified as a statistical outlier at the 1% level.

$$C = \frac{(N-1)t_{df-1,p}}{\sqrt{N\left(N-2+t^2_{df-1,p}\right)}}$$

where N is the sample size, t is the one-sided 100p percentage point from the t distribution with df the degrees of freedom associated with S:

$$p = 1 - \frac{0.01}{2N}$$

Alternative outlier methods are available that are intended for use on data sets that may contain multiple outliers and for detection of outliers associated with the bioassay design or model. For further discussion of outliers, see general information chapter *Analytical Data—Interpretation and Treatment* ⟨1010⟩, which may be a helpful, but not mandatory, resource.

THE CONFIDENCE INTERVAL AND LIMITS OF POTENCY

The following method (Fieller's) is used to determine the confidence interval for an estimate of log relative potency from a parallel line assay or a slope ratio assay. Let $M = a/b$ be the ratio for which we need a confidence interval. For the estimates, a and b, we have their respective standard errors, SE_a and SE_b, and a covariance between them, denoted Cov. The confidence interval, (M_{Low}, M_{Up}), for the estimated log relative potency then is as follows:

$$(M_{Low}, M_{Up}) = \left\{ \frac{M - \frac{gCov}{SE_b^2} \pm \frac{t}{b}\sqrt{\left(1-g\right)SE_a^2 + M^2SE_b^2 - 2MCov + \frac{gCov^2}{SE_b^2}}}{1-g} \right\}$$

where:

$$g = \frac{t^2SE_b^2}{b^2}$$

and $t = t_{df,\alpha/2}$ is the upper $\alpha/2$ percentage point (or the two-sided α percentage point) with the residual degrees of freedom, df, from the statistical analysis and chosen confidence level, 100*(1−α), (usually 95%). If $g \geq 1$, it means that the denominator, b, is not statistically significantly different from 0 and the use of the ratio is not sensible for those data. The length, L, of this confidence interval is $M_{Up} - M_{Low}$.

For those cases in which the estimates of a and b are statistically uncorrelated ($Cov = 0$), the confidence interval formula simplifies to the following:

$$(M_{Low}, M_{Up}) = \left\{ \frac{M \pm \frac{t}{b}\sqrt{\left(1-g\right)SE_a^2 + M^2SE_b^2}}{1-g} \right\}$$

For further discussion of confidence intervals for potency, see chapter ⟨1034⟩ which may be a helpful, but not mandatory, resource.

Change to read:

COMBINATION OF INDEPENDENT ASSAYS

When the monograph or chapter permits, multiple independent assays may be performed until the combined results reduce the confidence interval width to within the limits specified in the pertinent monograph or chapter. Where two or more independent assays are required, each leading to a log-potency M, the M's are combined using one of the following two methods.

Method 1

Let M_i denote the logarithm of the relative potency of the *i*th assay of *h* assay results to be combined. To combine the *h* results, the mean, standard deviation, and standard error of the M_i are calculated in the usual way:

$$\text{Mean } \overline{M} = \sum_{i=1}^{h} M_i / h$$

$$\text{Standard Deviation } S = \sqrt{\frac{1}{h-1} \sum_{i=1}^{h} \left(M_i - \overline{M}\right)^2}$$

$$\text{Standard Error } SE = S/\sqrt{h}$$

A 100(1 − α)% confidence interval is then found as:

$$\overline{M} \pm t_{h-1,\alpha/2} SE$$

where $t_{h-1,\alpha/2}$ is the upper α/2 percentage point (or the two-sided α percentage point) of a *t*-distribution with *h* − 1 degrees of freedom. The width, *L*, of this interval is $2t_{h-1,\alpha/2}SE$.

Method 2

It is assumed that the results of each of the *h* assays have been analyzed to give *h* values of log potency with associated confidence limits. For each assay, *i*, obtain the confidence interval for the log potency or log relative potency. Then compute value L_i by subtracting the *i*th lower confidence limit from the *i*th upper confidence limit. A weight w_i for each value of the log relative potency, M_i, is calculated as follows, where t_i has the same *t*-distribution value as that used in the calculation of confidence limits in the *i*th assay and is based on n_i degrees of freedom:

$$W_i = \frac{4t_i^2}{L_i^2}$$

The products $w_i M_i$ are formed for each assay, and their sum is divided by the total weight (*w*) for all assays to give the weighted mean log relative potency and its standard error as follows:

$$\left.\begin{aligned} \text{Mean, } \overline{M} &= \sum_{i=1}^{h} w_i M_i / w \\ \text{Approximate Standard Error, } SE &= 1/\sqrt{w} \\ \text{where } w &= \sum_{i=1}^{h} w_i \end{aligned}\right\} \quad (1)$$

Next compute an approximate chi-square:

$$\chi^2_M = \sum_{i=1}^{h} w_i \left(M_i - \overline{M}\right)^2 = \sum_{i=1}^{h} w_i M_i^2 - w\,\overline{M}^2$$

If the value of the approximate χ^2_M is *well under* the 5% value shown in Table 3, compute the confidence interval using the mean and approximate standard error equations in (1) above; otherwise use *Alternate weights* as described below. Labs need to specify in their procedures how to quantify "well under". Absent such a specification, the 20% values of Table 3 are suggested.

A 100(1 − α)% confidence interval in the log scale is then found as:

$$\overline{M} \pm L/2$$

$$\text{where } L = \frac{2t_{df,\alpha/2}}{\sqrt{w}} \sqrt{1 + \frac{4}{w^2} \times \sum_{i=1}^{h} \frac{w_i(w - w_i)}{n_i'}}$$

$$n_i' = n_i - 4 \times \frac{h-2}{h-1}$$

$$\text{and } df = w^2 \Big/ \sum_{i=1}^{h} w_i^2 / n_i$$

where $t_{N,\alpha/2}$ is the upper α/2 percentage point (or the two-sided α percentage point) of a *t*-distribution with degrees of freedom, *df*. The width of this interval is *L*.

Table 3. Critical Values for Approximate Chi-Square Test

	Critical Values	
h	**5%**	**20%**
2	3.841	1.642

	Critical Values	
h	**5%**	**20%**
3	5.991	3.219
4	7.815	4.642
5	9.488	5.989
6	11.070	7.289
7	12.592	8.558
8	14.067	9.803
9	15.507	11.030
10	16.919	12.242

Alternate weights: The observed variation among the estimated log potencies or relative potencies can be divided into two components:

- intra-assay variation for assay *i*:

$$V_i = 1/w_i$$

- inter-assay component of variation:

$$S_B^2 = \max\left\{0, \frac{1}{h-1}\sum_{i=1}^{h}\left(M_i - \overline{M}\right)^2 - \frac{1}{h}\sum_{i=1}^{h}V_i\right\}$$

▲▲ (ERR 1-Aug-2020)

For each assay, a weighting coefficient is then calculated as:

$$W_i' = \frac{1}{V_i + S_B^2}$$

The confidence interval is then found as:

$$\overline{M} \pm t \times SE',$$

$$\text{where } SE' = 1\sqrt{w'} \text{ and } w' = \sum_{i=1}^{h} w_i'$$

and *t*, the *t-distribution value*, is often approximated by the value 2.

For further discussion of combination of assays, see ⟨1034⟩, which may be a helpful, but not mandatory, resource.

APPENDIX—KEY LITERATURE

Bliss, CI. Analysis of the biological assays in *USP XV*. *Drug Standards*, 24, 33–67, 1956.

Böhrer A. One-sided and two-sided critical values for Dixon's outlier test for sample sizes up to *n* = 30. *Economic Quality Control* 23, 5–13, 2008.

Cochran, WG. The combination of estimates from different experiments. *Biometrics* 10, 101–129, 1954.

Fieller, EC. Some problems in interval estimation. *Journal of the Royal Statistical Society*, *B* 16, 175–185, 1954.

Iglewicz, B and Hoaglin, DC. *How to Detect and Handle Outliers*. Quality Press, Milwaukee, 1993.

Auxiliary Information - Please check for your question in the FAQs before contacting USP.

Topic/Question	Contact	Expert Committee
<111> DESIGN AND ANALYSIS OF BIOLOGICAL ASSAYS	Ravi Dasari Statistician	GCSTAT2020 General Chapters - Statistics

Most Recently Appeared In:
Pharmacopeial Forum: Volume No. PF 40(4)

Current DocID: GUID-E40D050D-B163-43F4-8D82-891AF3907355_3_en-US

Previous DocID: GUID-E40D050D-B163-43F4-8D82-891AF3907355_2_en-US
DOI: https://doi.org/10.31003/USPNF_M98860_03_01
DOI ref: h2qq0