

1 **Revision to USP General Chapter *Analysis of Biological Assays* <111>**
2 **(Fall 2007 version)**

3
4 **BRIEFING**

5
6 Although the analysis and design of bioassays has changed a great deal in the past 50 years, *United*
7 *States Pharmacopeia (USP)* General Chapter *Design and Analysis of Biological Assays* <111> has not
8 undergone significant revision. In 2001 the USP Biostatistics Expert Committee convened an Ad
9 Hoc Panel charged with the revision of <111>. That Panel has completed a draft revision, which
10 includes all topics, though lacks some planned examples. This revision will be published in
11 *Pharmacopeial Forum* as an In-Process Revision, and is also presented here.. The revised <111> is
12 intended to replace the current <111> in its entirety.

13
14 When completed, the revised <111> chapter, which will address analysis of biological assay data,
15 will be one of four in *USP* pertaining to bioassays. These chapters will entirely replace current
16 <111>. In addition to a “roadmap” chapter (as yet unnumbered) that will include a glossary
17 applicable to the other three chapters, there will be two new *USP* chapters: *Design and Development of*
18 *Biological Assays* <1032> and *Biological Assay Validation* <1033>. The draft glossary appeared in
19 *Pharmacopeial Forum* [2006;32(4):1359–1365].

20
21 The Panel encourages input from all interested parties regarding the revision of <111>.
22 Contemporary avenues of communication greatly expand the opportunity for involvement in
23 shaping the chapter. USP’s intent is to reflect the best contemporary thought regarding bioassay
24 analysis. This will be achieved when members of the bioassay community take advantage of this
25 opportunity to engage in the chapter’s development by responding to the material in this In-Process
26 Revision. Comments regarding this material should be sent to Larry N. Callahan, Ph.D.
27 (lnc@usp.org).

28
29 As a complement to the material that follows, USP plans to make available on its Web site data sets
30 that can be used by laboratories to verify commercial software packages. These data sets are
31 intended to be used to ensure that similarity is assessed correctly, that relative potency and the
32 associated confidence intervals are calculated correctly, and to assist in the proper selection of dose–
33 response models. Data sets will be developed for both linear and nonlinear response models. To
34 this end, the Panel invites the submission of data sets for consideration for use in this context.

35
36
37 **1.0 INTRODUCTION**

38
39 Although advances in chemical characterization have reduced the reliance on bioassays for many
40 products, bioassays are still essential for the determination of potency and the assurance of activity
41 of many proteins and complex mixtures, as well as for their central role in monitoring the stability of
42 biological products. The intended scope of revised General Chapter <111> includes guidance for
43 the analysis of results both of bioassays described in the *United States Pharmacopeia (USP)*, and of non-
44 *USP* bioassays that seek to conform to the qualities of bioassay design as recommended by *USP*.
45 (Note the emphasis on analysis: Design and validation are addressed in complementary chapters,
46 and an additional useful reference will include a Bioassay Glossary, which will be available as part of
47 a new General Chapter). Topics addressed in revised <111> include statistical concepts, and

48 methods of analysis for the calculation of potency and confidence intervals for a variety of bioassays,
49 including those referenced throughout *USP*. The revised <111> is intended primarily for use by
50 those who do not have extensive training or experience in statistics, but it also will provide guidance
51 for statisticians who are not experienced in the analysis of bioassays. In addition, the General
52 Chapter introduces selected modern methods, the implementation of which requires the guidance of
53 an experienced statistician.

54
55 The methods described are recommended but not mandatory; USP explicitly recognizes that
56 reasonable and sound alternatives may be employed. Additionally, Panel members assume that
57 computers and appropriate software will be used for data analysis. This latter view does not relieve
58 the bioassay practitioner of responsibility for the consequences of choices pertaining to bioassay
59 design and analysis.

60 61 2.0 CONCEPTS AND TERMINOLOGY

62
63 **NOTE:** *Some definitions included here are taken from the published glossary [Pharm Forum 32(4)]. This section*
64 *will be deleted from the final version of <111>.*

65
66 **Relative potency bioassay** – a comparison of biological responses on a continuous scale, measured
67 for Test and Standard materials, to derive a judgment regarding the capacity of Test material to
68 effect a specific response. The result from a relative potency assay quantitates the Test compound
69 capacity for achieving the specific response relative to the effect observed for the Standard
70 compound.

71
72 **Block, Blocking** – the grouping of related experimental units in experimental designs. Blocking is
73 often used to reduce the variability of a measure of interest. Blocks may consist of groups of animals
74 (a cage, a litter, or a shipment), individual 96-well plates, sections of 96-well plates, or whole 96-well
75 plates grouped by analyst, day, or batch of cells. The goal is to isolate a systemic effect, such as cage,
76 so that it does not obscure the effects of interest.

77
78 **Experimental design** – the structure of assigning treatments to experimental units. Blocking (q.v.),
79 randomization (q.v.), replication, and specific choice of design (to be covered in the planned General
80 Chapter *Design and Development of Biological Assays* <1032>) are some aspects of experimental design.
81 Important components of experimental design include the number of samples, the number of
82 concentrations, and how samples and concentrations are assigned to experimental units and are
83 grouped into blocks. The experimental design influences which statistical methodology should be
84 used to achieve the analytical objective.

85
86 **Experimental unit** – the smallest unit to which a distinct level of a treatment is randomly allocated.
87 An experimental unit needs to be distinguished from a sampling unit, the smallest unit on which a
88 distinct measurement is recorded (e.g., a well). Because the sampling unit is often smaller than the
89 experimental unit, it is an easy mistake to treat sampling units as if they are experimental units; this
90 mistake is called pseudoreplication. Different treatment factors can be applied to different
91 experimental units. For example, samples may be assigned to rows on a 96-well plate while dilutions
92 of the samples are assigned to columns on the plate. In this case, rows are the experimental units for
93 samples, and columns are the experimental units for sample dilutions.

94

95 **Independence** – For two measurements or observations A and B (raw data, assay sets, or relative
96 potencies) to be independent, values for A must be unaffected by B 's values and vice versa. A
97 consequence of non-recognition of lack of independence is poor characterization of variance. In
98 practice this means that if two potency or relative potency measurements share a common factor
99 that might influence assay outcome such as analyst, cell preparation, incubator, group of animals, or
100 aliquot of Standard sample, then the correct initial assumption is that these relative potency
101 measurements are not independent. As assay experience is gained an empirical basis may be
102 established so that it is reasonable to treat potency measures as independent even if they share a
103 common level of a factor. The same concern for lack of independence holds if two potency or
104 relative potency measurements are estimated together from the same model or are in any way
105 associated without including in the model some term that captures that there are two or more
106 potency measurements.

107
108 **Randomization** – a process of assignment of treatment to experimental units based on chance so
109 that all such experimental units have an equal chance of receiving a given treatment. The use of
110 randomization results in systematic error becoming random error not associated with particular
111 samples or a dilution pattern but distributed throughout the assay. In 96-well bioassays, plate effects
112 can be substantial and cause bias or trending, particularly in assays involving long-term cell culturing
113 or multiple addition and wash steps. In animal studies, a variety of factors associated with individual
114 animals can influence responses. If extraneous factors that influence either plate assays or animal
115 assays are not routinely demonstrated to have been eliminated or minimized so as to be negligible,
116 randomization that removes the influence of the biasing factor is essential to obtain unbiased data
117 required for the calculation of true potency. Randomization is central to the experimental design and
118 analysis of data associated with most biological assays.

119
120 **Reportable value** – the potency or relative potency estimate of record that is intended to achieve
121 such measurement accuracy and precision as are required for use. *Notes:* 1. The reportable value is
122 the value that will be compared to a product specification. The specification may be in the *USP*
123 monograph, or it may be set by the company, e.g., for product release. 2. The term *reportable value* is
124 inextricably linked to the “intended use” of an analytical procedure. Tests are performed on samples
125 in order to yield results that can be used to evaluate some parameter of the sample in some manner.
126 One type of test may be configured in two different ways because the resulting data will be used for
127 two different purposes (e.g., lot release vs. stability). The reportable value would likely be different
128 even if the mechanics of the test itself were identical. Validation is required to support the properties
129 of each type of reportable value. In practice there may be one physical document that is the
130 analytical procedure used for more than one application, but each application must be detailed
131 separately within that document. Alternatively, there may be two separate documents for the two
132 applications. 3. When the inherent variability of a biological response, or that of the log potency,
133 precludes a single assay data set's attaining a value sufficiently accurate and precise to meet an assay
134 specification, the assay, or analysis data set, may consist of multiple assay data sets, as necessary. The
135 number of assay data sets needed depends on the assay's accuracy and precision and on the intended
136 use and hence the properties of the reported value and is influenced by factors such as the type and
137 variability of the biological activity being studied.

138

3.0 A STEPWISE APPROACH TO ANALYSIS OF BIOASSAY DATA

Following is a set of steps that will help guide the analysis of a bioassay. Not all of the steps are needed for the analysis of every assay. Many of the steps (e.g., a choice of transformation or weighting scheme) may be addressed during development, checked during validation, and not repeated routinely. Details on some of the considerations are found in later sections, particularly Section 4.

- (1) Fit a means model (an analysis of variance model that fits a separate mean to each dilution level of each sample tested) with appropriate error terms.
- (2) During assay development, assess the distribution of the residuals, specifically examining them for departures from normality and constant variance, and transform the data as necessary; or if needed, choose a weighting scheme. Use as large a body of assay data as possible for this step—at least several assays, and preferably dozens of assays. The primary goal is to address any departure from constant variance. Step 2 can alternate between imposing a transformation and assessing the distribution of the residuals. After analysts have established a transformation and/or weighting scheme, it is not appropriate for operators to assess constant variance and normality for each instance of the assay. In addition to addressing transformation and weighting during assay development, analysts should reassess the choice of the transformation and weighting scheme if there have been changes to the assay that may influence precision or if system suitability parameters are trending in a manner that indicates the assay is or may become out of control.
- (3) Screen for outliers. This step normally follows the imposition of a suitable transformation or weighting method. Outlier analysis is best done on the residuals from a model that is fit to all the data within an assay. Outlier analysis on a small number of replicates or pseudoreplicates is not recommended. See Section 4.7 and *Analytical Data—Interpretation and Treatment* <1010>. In some cases outliers may be so severe that a reasonable model cannot be fit, and thus residuals will not be available. In such cases it is necessary to evaluate the raw data for outliers before one attempts to fit the model.
- (4) Remove outliers as appropriate. Statistical outlier tests should be seen as identifying points that are potential outliers. Before an observation is declared an outlier, investigations should routinely be conducted to determine whether a cause can be identified. Good practice will include recording the result of this investigation and the analyst who decides, based on his or her scientific judgment and the outlier test (if conducted) that an observation should be considered an outlier. Removing data as outliers should be rare. If many values from a run are removed as outliers, that run should be considered suspect. The assay procedure should state how many outliers are “acceptable.” Instructions for the investigation and treatment of an outlier observation should be included in the assay procedure and must be considered separate from the investigation and treatment of an out-of-specification (OOS) result (reportable value). Decisions to remove an outlier from data analysis should not be made on the basis of how the reportable value will be affected in terms of a potential OOS result.
- (5) Review the effectiveness of transformation (Section 4.4) after outlier removal.
- (6) Refit the means model and re-estimate error terms in the model.

- 187
188 (7) Choose subsets of data to use for linear model potency estimation and the model, whether
189 linear or nonlinear. This may mean identifying the linear range or the doses to use for fitting
190 a nonlinear model. In either case, assess whether the chosen model provides an adequate fit.
191

192 Analysts should not use the nonlinear regression fit to assess similarity or estimate potency if
193 either (a) inadequate asymptote information is available; or (b) a comparison of pooled error(s)
194 from nonlinear regression to pooled error(s) from a means model shows that the nonlinear
195 model does not fit well; or (c) other appropriate measures of goodness of fit show that the
196 nonlinear model is not appropriate (e.g., residual plots show evidence of a “hook”).
197

198 Additional specific guidance about selection of data subset(s) for linear model estimation of
199 relative potency include the following:

- 200 a. use at least three and preferably at least four to five adjacent concentrations
201 b. require that the slope is sufficiently steep
202 c. require that the fitted lines to Standard and Test samples be nearly straight
203 i. do not use difference testing—equivalence testing is recommended; see
204 Section 4.3
205 d. require that the fitted lines to Standard and Test samples be nearly parallel
206 i. do not use difference testing—equivalence testing is recommended; see
207 Section 4.2.
208
- 209 (8) For samples that yield similar dose–response curves (or similar straight-line subsets of
210 doses), calculate the relative potency estimate.
211
- 212 (9) If multiple independent assays (using independent samples) of the same analyte are available,
213 construct a variance estimate and confidence interval for the estimated potency of the
214 analyte using the sampling variance of the log potency from the independent assays (as
215 described in Section 6.2.1). Use a Fieller’s Theorem–based confidence interval for the
216 estimated potency only if no such sampling-based confidence interval is available. See
217 Section 4.10.
218

219 4.0 ELEMENTS OF DATA ANALYSIS

220
221 The elements of data analysis elements discussed below include essential biological and statistical
222 considerations that pertain to effective utilization of bioassay data in the calculation of relative
223 potency.
224

225 **4.1 Assumptions.** The key assumption for the analysis of most bioassays is that the Standard and
226 Test samples contain the same active analyte or population of analytes and nothing else that
227 differentially affects response. The existence of a Standard–Test sample pair that passes the
228 assessment of similarity is a necessary but not sufficient test that this key assumption is satisfied
229 (thus, it remains an assumption). In many assays multiple compounds will yield similar
230 concentration–response curves. Although it may be reasonable to use a biological assay system to
231 describe or even compare response curves from different compounds, it is not appropriate to report
232 relative potency unless the Standard and Test samples contain only the same active analyte.
233 Biological products typically exhibit lot-to-lot variation in the distribution of analytes. Similarity is

234 then, at least partially, an assessment of whether the distribution of analytes in the Test sample is
235 close enough to that of the Standard sample for relative potency to be meaningful.

236
237 An additional statistical assumption is that the design structure of the assay is reflected in the
238 analysis. Note that a design structure implies randomization and attention to experimental units and
239 (if present) blocks; see <1032>. If additional statistical assumptions are satisfied, then a simple
240 statistical analysis is reasonable. These assumptions are:

- 241
242 (1) independence of the observations,
243
244 (2) constant variance of the responses around the fitted model (may require an appropriate data
245 transformation), and
246
247 (3) normally distributed residuals—again, possibly after an appropriate transformation.
248

249 Bioassays routinely performed with cell cultures (using 96-well plates and multichannel pipettes)
250 commonly violate one or more of these assumptions. Wells prepared using serial dilution or using a
251 multichannel pipette are not independent. Addressing this lack of independence requires either a
252 change in laboratory practice or a more complex statistical analysis. Most statistical methods for
253 analysis of bioassay data are reasonably robust to mild departures from assumptions regarding
254 constant variance or normality. In many assay systems a good choice of transformation often will
255 yield bioassay data that are sufficiently close to constant variance and normality. However, methods
256 of statistical analysis are generally not robust to lack of independence. Therefore, it is essential that
257 the statistical model should capture the design structure imposed by the presence of blocks or
258 grouping of responses (e.g., due to serial dilution or use of a multichannel pipette). However, where
259 assay results are combined the independence of assays is of critical importance. Failure to fully
260 account for the lack of independence within an assay means that the estimates will not be as precise
261 as they could be, but they still may be useful. Also see the discussion of sample-based confidence
262 intervals in Section 4.10.

263
264 **4.2 Parallelism.** Many procedures for assessing parallelism are based on a statistical null hypothesis
265 of parallelism and an alternative hypothesis of some measurable nonparallelism. Failure to find
266 statistically significant nonparallelism is then taken as a conclusion of parallelism. This is neither a
267 proper conclusion nor an acceptable approach. Rather, to support a conclusion of parallelism
268 (*similarity*), analysts should use equivalence testing wherein “sufficiently parallel” (only trivial
269 departure from parallelism) is the alternative statistical hypothesis, and nonparallelism is the null
270 hypothesis. Some care should be taken with this approach: Departures from parallelism that are
271 consistent in value across many assays could be indicative of matrix effects or of real differences
272 between Test and Standard materials even if the nonparallelism is small. The steps for determining
273 parallelism for one such approach are:

- 274
275 (1) Choose a measure of nonparallelism. For the linear case, this could be the difference or ratio
276 of slopes. (The ratio of slopes can be less sensitive to the value of the slope. Also, framing
277 the slope difference as a proportional change from Standard rather than in absolute slope
278 units has an advantage because it is invariant to the units on the dose axis.) For the four-
279 parameter logistic model, similarity between Standard and Test samples must be assessed on
280 the basis of three parameters: the upper asymptote, the slope, and the lower asymptote. If a

281 five-parameter logistic curve is used to fit bioassay data it would also be important to show
282 that the fifth parameter of the curve also is similar between Standard and Test preparations.
283

284 (2) Specify a range of acceptable values, typically termed an equivalence interval or “indifference
285 zone,” for the measure of nonparallelism. When a ratio of slopes is used as a measure of
286 nonparallelism, that measure is free to vary above or below a ratio of 1.0 (a ratio of 1.0
287 indicates perfect parallelism). The acceptable values for nonparallelism will then constitute
288 an interval.
289

- 290 a. Two approaches can be used to determine this interval. The first is to base it on what
291 is known of the product and the assay, much as (0.80, 1.25) is used for
292 bioequivalence of generic products, or based on general experience (see Examples
293 7.1 and 7.3). Considerations here could include the therapeutic index of the drug and
294 the precision required from the assay.
295 b. The second approach is to compile historical data that compare the Standard to itself
296 and determine, from the historical data, the equivalence interval as a tolerance
297 interval for the measure(s) of nonparallelism. The advantage of using historical data
298 is that they give the laboratory control of the false failure rate (the rate of failing an
299 assay that is in fact acceptable). The disadvantage is that there is no control of the
300 false pass rate (the rate of passing an assay that is not working right). The
301 equivalence interval specification will be driven solely by assay capability.
302 Laboratories that use this approach need to take care that an assay in need of
303 improvement is not driving overly wide equivalence intervals. Also note that any
304 change in assay capability means changing the equivalence interval. The interval
305 needs to be relevant to the assay as it is currently being conducted.
306

307 (3) Determine a 90% confidence interval for the measure of nonparallelism. If this confidence
308 interval lies entirely within the equivalence interval specified in Step 2, then similarity is
309 sufficiently demonstrated.
310

311 An alternative to the approach described above is to use an average (historical) value for the variance
312 of the ratio or difference in a similarity parameter—obtained from some number of individual
313 assays—to compute an acceptance interval for a point estimate of the similarity parameter. This
314 approach is simpler to implement in routine use and can be used with assay designs that are unable
315 to provide reliable estimates of within-assay variation. However, there is a price. The previously
316 described equivalence testing approach that relies on assay-specific (within-assay) measure(s) of
317 variation (i.e., the confidence intervals) is conservative in the sense that it will fail to pass similarity
318 for samples from assays that have larger than usual amounts of within-assay variation. Using an
319 acceptance region for a similarity parameter—rather than an acceptance region for confidence
320 intervals for the similarity parameter—loses this conservative property and hence is not preferred
321 where alternatives exist.
322

323 **4.3 Linearity of Concentration–Response Data.** Some analyses of bioassay assume that the shape
324 of the concentration–response curve is a straight line or approximates a straight line over a limited
325 range of concentrations. In those cases, it is appropriate to assess whether a linear response model is
326 justified for the data in hand. Standard methods for assessing linearity face the same problems as do
327 standard methods for assessing parallelism: Specifically, more data and better precision make it more
328 likely to detect nonlinearity, and the tests for nonlinearity are typically set up to confirm the null

329 hypothesis. Because the situations when nonlinearity does not affect the potency are generally rare
330 and may occur only by chance, analysts should routinely assess departure from linearity if they wish
331 to use the linear response model to estimate potency.

332
333 The general approach for linearity should follow that for similarity. One needs to specify a measure
334 of departure from linearity, identify a limit for that measure that specifies acceptable values of
335 nonlinearity and then perform an appropriate statistical equivalence test. A procedure may take the
336 following form:

- 337
338 (1) If an examination of a plot of the data reveals gross departure from linearity, this is sufficient
339 to support a conclusion that linearity is not present. More formally,
340 (2) Construct 90% confidence intervals for observations of a measure of curvature made
341 following assay optimization using historical data from the assay run under conditions that
342 are similar to current conditions *and for which linearity is accepted on nonstatistical grounds*.
343 (3) Specify a symmetric equivalence interval that captures approximately 95% to 99% of these
344 intervals.

345
346 **4.4 Transformations.** Transformations are used primarily to address nonconstant variation in the
347 response and normality of residuals and secondarily to improve the fit of a statistical model to the
348 data. Bioassay data are commonly displayed with concentration on a log scale (in parallel-line or
349 parallel curve analyses). Slope ratio assays are displayed with concentration on the original scale. This
350 section concerns transformations that are a re-scaling of the response axis, not of the concentration
351 axis. Common choices for a transformation of the response include log, square root, reciprocal, and,
352 for count data with known asymptotes, logit. The log and reciprocal transformations share the
353 advantage of easy transformation back to the original scale.

354
355 From a statistical perspective there is nothing special about the original scale of measurement. Any
356 transformation that resolves a nonconstant variance is acceptable. Analysts should recognize,
357 however, that transformations may make interpretation difficult. Log transformations are commonly
358 used, partly because of the ease of transforming back to the original scale for interpretation. Also, if
359 a transformation is used, then that affects the curve fit. The reverse may also be a consideration.
360 That is, by transforming the response, e.g., by a log or square root, changes the shape of the
361 response curve and may change the range of doses for which the responses are nearly straight and
362 nearly parallel. For example, for an assay with an untransformed response that indicates variance
363 increasing with response and parallel straight lines, after transformation the responses will no longer
364 fit a straight-line model well.

365
366 In many bioassays a suitable transformation (often the logarithm) yields data that have nearly
367 constant variance, near normal distribution of the residuals, and a good fit to a four-parameter
368 logistic curve. When no single transformation yields all these properties a more complex approach,
369 such as weighting, is called for. It is particularly appropriate to seek competent statistical guidance
370 for the process of choosing a transformation and/or weighting scheme for the analysis of a bioassay.
371 The choice of transformation (and if necessary, a weighting method) for a bioassay is normally made
372 prior to validation. It is not appropriate to choose a different transformation each time the assay is
373 run.

374
375 **4.5 Normality.** Many statistical methods for the analysis of quantitative responses assume normality
376 of the residuals. If the normality assumption is not valid, the estimate of relative potency and its

377 standard error may be valid, but a confidence interval for the relative potency estimate will not be
378 valid. Most methods are reasonably robust to departures from normality, so the goal here is to detect
379 substantial non-normality. The primary tool is a normal probability plot and a histogram (or
380 something similar, like stem-and-leaf or box plots) of the residuals from the analysis. The histogram
381 should appear unimodal and symmetric. The normal probability plot should follow approximately a
382 straight line. An S-shaped pattern about a straight line is one indication of non-normality. This
383 assessment typically is performed during assay development and/or development and is not
384 routinely conducted with each assay run.

385
386 **4.6 Variance Heterogeneity.** As mentioned in Section 4.1, one common assumption for analyses
387 with quantitative responses is that of equal variances among the responses, which is also referred to
388 as homogeneity of variance. For most common linear and nonlinear regression models, the variance
389 referred to here is the residual variance. If the variances are not all equal but the data are analyzed as
390 if they are all equal, then the estimate of relative potency may still be reasonable but it may be less
391 precise than it could be. More importantly, the assessment of similarity will not be appropriate and
392 standard errors and confidence intervals for all parameters (including a Fieller's Theorem-based
393 interval for the relative potency) reported from the analysis will not be correct and should not be
394 used.

395
396 *Weighting* is an approach to dealing with unequal variances. (If the coefficient of variation is constant,
397 then the variance is not; see Example 7.3.) Basically, each observation should be weighted by a
398 quantity that is proportional to the reciprocal of the variance of that observation.

399
400 Variance may be proportional to dose or some function of dose. This possibility may be examined
401 by plotting the sample variance at each dose against dose and then against a function of dose (e.g.,
402 dose squared). Variance will be proportional to the function of dose where the plot follows
403 approximately a straight line. (If the data fall along a horizontal line, no weighting is needed.) If such
404 a function can be found, then the weights are taken as the reciprocal of that function. There may be
405 no such function or at least not one that is easy to describe, particularly if the variation is higher at
406 both extremes of the concentration range studied. An alternative is to develop an estimate of the
407 variability at each concentration using historical data for the assay as the latter is currently
408 conducted. Using only the replication variances from the current assay is not appropriate. There are
409 too few data to properly determine truly representative variances specific to each dose. Whether one
410 uses a model or historical data, the goal is to capture the relative variability at each concentration.
411 Even with use of historical data, one does not necessarily have to assume that the absolute level of
412 variability of the current assay is identical to that of the historical data, but only that the ratios of
413 variances among dilutions are the same.

414
415 There is no substitute for training and experience in determining an appropriate variance modeling
416 strategy. For example, data that have constant variation around a four-parameter logistic dose-
417 response curve but have appreciable variation in the effective dilution 50% (ED50) parameter from
418 block to block within the assay or from assay to assay can easily appear to have large variation in the
419 response for doses near the long-term average value of ED50. A weighted model with low weights
420 for doses near the ED50 would quite badly misrepresent a major feature of such an assay system.

421
422 **4.7 Outliers.** An outlier is a datum that appears not to belong among the other data present. An
423 outlier may have a distinct, identifiable cause, such as a mistake in the bench work, equipment
424 malfunction, or a data recording error, or it could just be an unusual value relative to the variability

425 typically seen and may appear without an identifiable cause. The essential question pertaining to an
426 outlier becomes: Is the apparent outlier sampled from the same population as the other, less
427 discordant, data, or is it from another population? If the former is the case and the datum is an
428 unusual (yet still legitimate) value obtained by chance, then the datum stands and should be retained
429 in data analysis; if the latter is the case and the datum's excursive value is due to human error or
430 instrument malfunction, then the datum may be omitted from calculations. Outlier management
431 relies on procedures and practices to yield the best answer possible to that essential question and to
432 respond accordingly.

433

434 General Chapter *Analytical Data – Interpretation and Treatment* <1010> addresses outlier labeling,
435 identification, and rejection, including statistical methods and provides material that the bioassay
436 practitioner will find useful. General Chapter <1010> also lists additional sources of information
437 that can provide a comprehensive review of the relevant statistical methodology. Some remarks
438 about outliers are provided here in the context of bioassays to emphasize or complement what is
439 found in <1010>.

440

441 Of the procedures employed for analysis of drug compounds and biological moieties, the bioassay
442 can be expected to be the most prone to outlying data. The management of outliers is appropriate
443 with bioassay data on at least two levels: 1) individual responses can be checked against expected
444 responses for the sample and concentration; and 2) separately, estimates of relative potency from an
445 assay can be checked for consistency with other independent estimates of the potency of the same
446 material. The following four steps may be considered in the bioassay control of outlying data.

447

448 1. Prevention—The first consideration is to develop procedures that are less subject to error and to
449 have in place checks that are sensitive to the sorts of errors that, given the experience in assay
450 development, may be expected to occur. In effect, the error never becomes an outlier because it is
451 identified as it occurs.

452

453 2. Labeling—It is good practice to examine data for the outlier and flag (“label”) the apparently
454 outlying observation for investigation. If investigation finds a cause, then the outlying datum may be
455 excluded from analysis. Given the ordinary occurrence of substantial variability in bioassay response,
456 a laboratory's investigation into whether a cause for the outlying observation can be identified is
457 likely to yield no determinable cause. Yet the lack of evidence regarding an outlying observation's
458 cause is not a clear indication that statistical outlier testing is warranted. Knowledge of the typical
459 range of assay response variability should be the foundation for a justification for the use of
460 statistical outlier tests.

461

462 3. Identification—Outlier identification is the use of significance tests to confirm that the values are
463 inconsistent with the known or assumed statistical model. For outliers with no determined cause, it
464 is tempting to use statistical outlier identification procedures to discard unusual values from the
465 analyses. Discarding data solely because of statistical considerations should be a rare event. Falsely
466 discarding data leads to overly optimistic estimates of variability and can bias potency estimates.

467

468 Statistical procedures depend on assumptions about distribution. Identification of data as outliers
469 may mean only that the assumption about distribution is not correct. If dropping outliers because of
470 statistical considerations is common, particularly if outliers tend to occur more often at high values
471 or at high responses, then consideration needs be given to the assay procedures and to the
472 assumptions underlying the statistical method.

473
474 General Chapter <1010> makes no explicit remarks regarding outlier analysis in linear or nonlinear
475 regression. Presuming that it is deemed appropriate for statistical analysis of data for the presence of
476 outliers to proceed, techniques specific to regression may profitably be employed when the bioassay
477 is comprised of regressions of response on dose.

478
479 4. Robust methods—Robust methods are less sensitive to influence by outlying observations. An
480 alternative to discarding outlying data is to use methods for data description and statistical analysis
481 that are less affected by the presence of the outlier(s), e.g., using the median rather than the mean to
482 describe the data's center. Also, regression using the method of least squares, which underlies many
483 of the methods in this Chapter, is not robust to the presence of outliers. The use of contemporary
484 methods of robust regression may be appropriate.

485
486 **4.8 Replacement of Missing Values.** Unless directed by a specific monograph, analysts generally
487 do not need to replace missing values. Most modern statistical methodology (and software) does not
488 require equal numbers at each combination of dose and product.

489
490 **4.9 Pooling Variance/Variance Components.** Pool individual variance estimates wherever
491 possible, e.g., between blocks. Even with modest replication there is considerable gain in
492 characterization precision in pooling variation. It is important to assess whether or not it is
493 reasonable to pool data before actually doing so. If this assessment is done during assay
494 development and validation, it may not be necessary or appropriate in routine use of the assay.

495
496 **4.10 Confidence Intervals.** A report of an assay result should include a measure of the uncertainty
497 of that result. This can be a standard error or a confidence interval. An interval (c, d) is a 95%
498 confidence interval for a parameter (e.g., relative potency) if 95% of such intervals upon repetition
499 of the experiment would include the actual value of the parameter. One can think of a confidence
500 interval as indicating values of the parameter that are consistent with the data.

501
502 Confidence intervals can be determined by combining results for the logarithm of the relative
503 potency from independent assays as described in Section 6.2.1. Confidence intervals can either be
504 *sample based* or *model based*. A model-based interval is based on the standard errors for each of the one
505 or more estimates of log relative potency that come from the analysis of a particular statistical
506 model. Model-based intervals should be avoided if sample-based intervals are possible. Model-based
507 intervals require that the statistical model correctly incorporate effects and correlations that
508 influence the model's estimate of precision. These include but are not be limited to serial dilution
509 and plate effects. Sample-based methods require only independence and that the independent assays
510 be conducted in the same manner. Also, many assay procedures fix the acceptable width of the
511 confidence interval, and two or more independent assays may be needed to meet the specified limit.
512 This discussion should not be construed to dismiss the value in addressing correlations and other
513 factors that influence within-assay precision to the extent possible. The within-assay precision is a
514 portion of the variability that is the basis for the sample-based intervals, and thus minimizing it to
515 the extent possible is still valuable. This section describes Fieller's Theorem, a commonly used
516 model-based interval. Sample-based intervals are covered in Section 6.2.1.

517
518 Many confidence intervals are of the form:
519

520 Confidence interval = value $\pm k \times$ the standard error of that value.

521
 522 For such cases as long as the multiplier k can be easily determined (such as from a table of the t -
 523 distribution), reporting the standard error and the confidence interval are largely equivalent because
 524 the confidence interval is then easily determined from the standard error. For ratios such as the
 525 logarithm of relative potency, the confidence interval is not symmetric around the point estimate;
 526 Fieller's Theorem is needed. For these asymmetric cases the confidence interval should be reported
 527 because the standard error by itself does not capture the asymmetry.

528
 529 Fieller's Theorem is the formula for the confidence interval for a ratio. Note that the relative
 530 potency from either a slope ratio or parallel-line bioassay fit by straight lines will be a ratio. Also
 531 note that for some parameterizations of nonlinear models, the relative potency will also be a ratio.

532 Let $R = a/b$ be the ratio for which we need a confidence interval. For the estimates of a and b we
 533 have their respective standard errors, SE_a and SE_b , and a covariance between them, denoted CO .
 534 (The covariance is a measure of the degree to which a and b are related and is proportional to the
 535 correlation between a and b .) The covariance may be 0, such as for standard linear (parallel-line)
 536 analyses, but it need not be. The confidence interval for R then is as follows, where the $\hat{}$ symbol
 537 indicates estimate.

$$(R_L, R_U) = \frac{\left\{ \hat{R} - \frac{gCO}{SE_b^2} \pm \frac{t}{\hat{b}} \sqrt{(1-g)SE_a^2 + \hat{R}^2 SE_b^2 - 2\hat{R}CO + \frac{gCO^2}{SE_b^2}} \right\}}{1-g}$$

538
 539 where

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

540
 541 and t is the appropriate t deviate value that will depend on the sample size and confidence level
 542 chosen (usually 95%). If $g > 1$, it means that the denominator, \hat{b} , is not statistically significantly
 543 different from 0 and the use of the ratio is not sensible for those data.

544
 545 For those cases where the estimates of a and b are statistically uncorrelated ($CO = 0$), the confidence
 546 interval formula simplifies to

$$(R_L, R_U) = \frac{\left\{ \hat{R} \pm \frac{t}{\hat{b}} \sqrt{(1-g)SE_a^2 + \hat{R}^2 SE_b^2} \right\}}{1-g}$$

547
 548 See Examples 7.1 and 7.2 for examples of use of Fieller's Theorem.

5.0 ANALYSIS MODELS

549
 550 A number of mathematical functions can be successfully used to describe a dose-response
 551 relationship. The first consideration in choosing a model is the form of the assay response. Is it a
 552 number, a count, or a category such as dead/alive? The form will identify the possible models that
 553 can be considered.
 554
 555
 556

557 Once the form of the response is identified, the next step is to select an appropriate dose–response
558 model. This step requires specific data. Four-parameter and five-parameter logistic equations of the
559 ln dose vs. response are often used, but other models may be more suitable. (See section 5.3).
560

561 **5.1 Quantitative and Qualitative Assay Responses.** The terms *quantitative* and *qualitative* refer to
562 the nature of the response of the assay used in constructing the dose–response model. Assays with
563 either quantitative or qualitative responses can be used to quantify product potency. Note that the
564 responses of the assay at the doses measured are not the reportable value (relative potency) of the
565 bioassay. It is important to understand the differences among the responses, dose–response
566 functions, and reportable values.
567

568 A quantitative response results in a number on a continuous scale. Common examples include
569 spectrophotometric and luminescence responses, body weights and measurements, and data
570 calculated relative to a standard curve (e.g., cytokine concentration). Models for quantitative
571 responses can be linear or nonlinear; see Section 5.3.
572

573 A qualitative measurement results in a categorical response. For bioassay, qualitative responses are
574 most often quantal, meaning they entail two possible categories such as Pass/Fail,
575 Positive/Negative, 0/1, or Dead/Alive. Quantal responses may be reported as proportions (e.g., the
576 proportion of animals in a group displaying a property). Quantal models are presented in Section
577 5.4. Qualitative responses can have more than two possible categories, with or without ordering of
578 the categories. These models are not considered in this General Chapter.
579

580 Assay responses can also be counts, such as number of plaques or colonies. Integer responses are
581 sometimes treated as quantitative, sometimes as qualitative, and sometimes models specific to
582 integers are used. The choice is often made based on the range of counts. If the count is mostly 0
583 and rarely greater than 1, the assay may be analyzed as quantal and the response is Any/None. If the
584 counts are large and cover a wide range, such as 500 to 2500, then the assay may be analyzed as
585 quantitative. A square root transformation of the count is often helpful in such analyses to better
586 satisfy homogeneity of variances. If the range of counts includes or is near 0 but 0 is not the
587 preponderant value, it may be preferable to use a model specific for integer responses. Poisson
588 regression and negative binomial regression models are often good options.
589

590 Integers are not the only responses that provide flexibility in the form of model. Assays with
591 quantitative responses may be converted to quantal responses. For example, what may matter is
592 whether some defined threshold is exceeded. The model could then be quantal—threshold
593 exceeded—or not. In general, assay systems will have better precision of potency if the model uses
594 all the information in the response. Using above or below a threshold, rather than all the measured
595 responses, is likely to degrade the performance of an assay.
596

597 **5.2 Fixed and Random Effects in Models of Bioassay Response.** The choice of treating design
598 factors as *fixed* or *random* is important both to the design and statistical analysis of the assay. We
599 consider this choice here with a focus on analysis; also see <1032> for design considerations.
600

601 Fixed effects are factors where all levels that exist or all in which there is interest in evaluating them
602 are discretely present, e.g., dose. Fixed effects are expected to cause a consistent shift in responses.
603 Some examples of fixed effects include temperature and time of thaw. Analysts study fixed effects
604 by controlling them in the design and examining changes in means across levels of the factor. In a

605 bioassay, sample and dose are fixed effects. Studies of other fixed effects are typically performed
606 early in the assay development cycle in the process of optimizing assay performance; see <1032>.

607
608 Random effects are factors where the levels in a particular run of an assay are considered
609 representative of levels that could have been present. That is, there is no expectation of a particular
610 value for a random effect. Rather, that value may vary subject to some expected distribution of
611 values and thus may be a source of variability. Random effects may be exemplified by reagent lot,
612 operator, date of assay, and well position on a plate, if there is no interest in *specific* reagent lots,
613 operators, dates, or plate well as sources of variability. Analysts study random effects by measuring
614 variance components. As with fixed effects, these are studied during assay development to
615 understand their influence on assay precision and ruggedness; see <1032>.

616
617 The choice of treating a factor as fixed or random is important to the design of the assay and to
618 proper reporting of the precision of the assay. Treating all factors as fixed, for example, leads to an
619 understatement of assay variability because it ignores all sources of variability other than replication.
620 The goal is to identify specific sources of variability that can be controlled, to properly include those
621 factors in the design, and then to include other factors as random.

622
623 In addition, if the factor switches or may switch from random to fixed effect or vice versa the factor
624 should be modeled as a random effect. For example, reagent lots cannot be controlled, so different
625 lots are typically considered to cause variability, and reagent lot would be considered a random
626 effect. However, if a large shift in response values has been traced to a particular lot, a comparison
627 among a set of lots could be performed using the predicted levels of each lot's random effect

628
629 Assay designs that consist of multiple factors are efficient, but they require corresponding statistical
630 techniques that incorporate the factors as fixed or random effects in the analysis. If all factors are
631 fixed, the statistical model is termed a *fixed-effects model*. If all are random, it is termed a *random-effects*
632 *model*. If some factors are fixed and some random, the model is a *mixed-effects model*. Note that the
633 concepts of fixed and random effects apply to qualitative and integer models as well as quantitative
634 models.

635
636 For assay designs that include multiple experiment units (i.e., samples assigned to sets of tubes and
637 doses assigned to pre-plate tubes) a mixed model in which the experimental units treated are as
638 random effects is a particularly effective way to approach the analysis.

639
640 **5.3 Linear and Nonlinear Models for Quantitative Responses.** Quantitative assays are
641 characterized by the assay measurement's being a number on a continuous scale. Optical density
642 values from immunological plate-based assays are such measurements. Models for quantitative
643 assays can be linear or nonlinear. While there is an apparent difference in level of complexity, there
644 is also much commonality among the parallel-line (linear) and parallel curve (nonlinear) models.
645 Because of the different form of the equations, slope ratio assays are considered separately (Section
646 5.3.4).

647
648 **5.3.1 Assumptions—parallel lines and curves.** The basic parallel-line and parallel-curve models
649 share some assumptions. Both include a residual term, ϵ , which is assumed to be independent and to
650 have constant variance from dose to dose. Often the residual term is assumed to have a normal
651 distribution as well. The assumptions of independence and equal variances are commonly violated,

652 so the goal in analysis is to incorporate the lack of independence and the unequal variances into the
653 statistical model or the method of estimation.

654
655 Lack of independence often arises because of the design or conduct of the assay. For example, if the
656 assay consists of responses from multiple plates, observations from the same plate are likely to share
657 some common influence that is not shared with observations of other plates. This is an example of
658 intraplate correlation. A simple approach for dealing with this lack of independence is to include a
659 term in the statistical model for plate. With enough plates this should be a random effects term so
660 that we obtain an estimate of plate-to-plate variability.

661
662 In general, the model needs to closely reflect the design. The basic model equations apply only to
663 completely randomized designs. Any other design will mean additional terms in the statistical model.
664 More examples are provided with the Examples (Section 7).

665
666 **5.3.2 Calculation of potency.** A primary assumption underlying methods used for the calculation
667 of relative potency is that the Test preparation behaves as a dilution (or concentration) of the
668 standard preparation. This condition is known as *similarity*. Similarity can be represented
669 mathematically as follows. Let F_T be the concentration–response curve for the Test, and let F_R be
670 the concentration–response curve for the Standard. The underlying mathematical model for
671 similarity is:

$$672 \quad F_T(\bar{z}) = F_R(\rho \bar{z}), \quad [5.1]$$

673 where \bar{z} represents the dose or concentration and ρ represents the relative potency of the Test
674 sample relative to the Standard sample.

675
676 Methods for estimating ρ in some common dose–response models are discussed below. For linear
677 models the distinction between parallel-line models (Section 5.3.2.3) and slope-ratio models (Section
678 5.3.4) is based on what transformation of \bar{z} (the dose) makes the model nearly linear over the range
679 of interest. That is, let $x = \bar{z}^\lambda$ for $\lambda \neq 0$, or $x = \log(\bar{z})$ for $\lambda=0$. If the model is linear in $\log(\bar{z})$, the
680 model is a parallel-line model. If the model is linear in \bar{z}^λ for any $\lambda \neq 0$, the model is a slope-ratio
681 model. For nonlinear models convention is to use $\log(\bar{z})$, though there is no theoretical reason
682 against an alternative transformation of \bar{z} .

683
684 **5.3.3 Linear models.** In this section, a linear model refers to a linear dose response, which is a
685 straight-line function between the dose, X , and the response, Y . Y may be the response or a
686 transformation of the response. The functional form of this relationship is $Y = a + bX$. Often, the
687 dose must be log transformed to result in a straight line. Then, X equals $\log(\text{dose})$. Straight-line fits
688 may be used for portions of a nonlinear dose–response curve, though doing so requires a method
689 for selecting the doses to use for each of Standard and Test samples.

690
691 **5.3.3.1 Means models vs. regression.** A linear dose–response model is most often analyzed with
692 ordinary least squares regression. Such an analysis results in estimates of the unknown coefficients
693 (intercepts and slope) and their standard errors, as well as measures of the goodness of fit [e.g., R^2 or
694 root mean square error (RMSE)].

695
696 Linear regression works best where all doses can be used and there is negligible curvature in dose–
697 response data. Another statistical method for analyzing linear dose–response curves is the *means*
698 *model*. This is an analysis of variance (ANOVA) method that offers some advantages. A means

699 model that makes use of contrasts will better capture the variance in the bioassay system than will
700 regression.

701
702 **5.3.3.2 Selecting linear range.** Often a subset of the doses measured in the assay must be selected
703 in order to create a linear dose–response curve. The subset can often be identified graphically
704 following a ranging study. It is important to choose a linear range that will result in straight lines for
705 the range of relative potencies expected during routine use of the assay. Otherwise the assay will fail
706 parallelism tests (see Section 4.2) when the problem is the potency, resulting in values outside the
707 linear range and entailing repeat assays.

708
709 The problem is more complex in assays where there is even modest variation in the shape or
710 location of the dose–response curve from assay to assay or from block to block within the assay. In
711 such assays, which are quite common, it is appropriate to choose subsets for each sample in each
712 assay or even in each block within an assay. Note that a fixed-effect model will mask any need for
713 different subsets in different blocks, but a mixed model can reveal and accommodate different
714 subsets in different blocks.

715
716 An assay with a linear dose–response curve can probably be shown to have an “s” shape if it is
717 tested over a wide enough range because plateaus of effect will be observed on the low- and high-
718 dose ends. Because the optical density (OD) is linear from about 0.1 to 2.0 Absorption Units (AU)
719 for many instruments, an asymptote near or outside 0.1 or 2.0 may be suspect as an artifact of the
720 instrument rather than an asymptote of the biological response.

721
722 **5.3.3.3 Parallel-line dose–response models.** If the concentration–response model (5.1) can be
723 made linear in $x = \log(z)$, the resulting equations are then:

$$724 \quad y_R = \alpha + \beta \log(z) + e = \alpha + \beta x + e$$

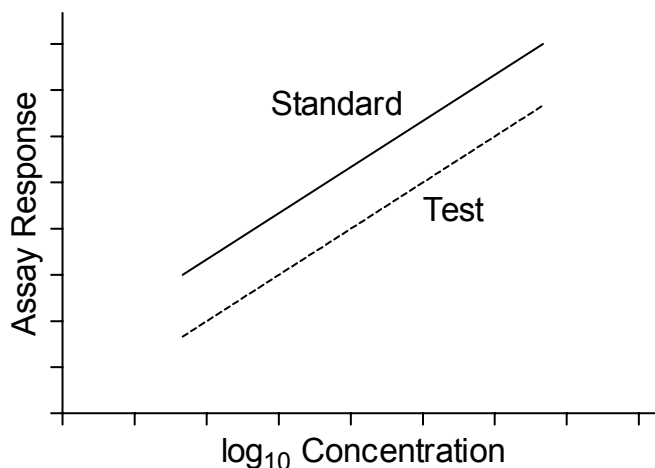
$$725 \quad y_T = \alpha + \beta \log(\rho z) + e = [\alpha + \beta \log(\rho)] + \beta x + e \quad [5.2]$$

726
727 where R denotes (Reference) Standard and T denotes Test. Where dose–response lines are parallel, a
728 separation or horizontal shift in one line indicates a difference in the level of biological activity being
729 assayed. This difference is quantitated by the difference in logarithms or log ratio of relative potency.

730

731

Figure 5.1. Example of parallel-linear model

732
733

734 Let α_R represent the x -intercept for the Standard, α_T the x -intercept for the Test, and β the common
735 slope. Then the relative potency is

736

$$\rho = \exp\left(\frac{\alpha_T - \alpha_R}{\beta}\right)$$

737

738 **5.3.3.4 Estimation of Parallel-line Models.** Parallel-line models are fit by the method of least
739 squares. If the equal variance and independence assumptions hold, the parameters of [5.2] are
740 chosen to minimize

$$741 \quad \sum (y - \hat{\alpha} - \hat{\beta}_1 T - \hat{\beta}_2 x)^2 \quad [5.3]$$

742 where the carets denote estimates. This is a linear regression with two independent variables, T and
743 x , where T is a variable that equals 1 for observations from the Test and 0 for observations from the
744 Standard. The summation in [5.3] is over all observations of the Test and Standard. If the equal
745 variance assumption does not hold but the variance is known to be proportional to a value, w , that
746 can be determined for each observation (see Section 4.6), then the method is weighted least squares

$$747 \quad \sum \frac{1}{w} (y - \hat{\alpha} - \hat{\beta}_1 T - \hat{\beta}_2 x)^2 \quad [5.4]$$

748 In both [5.3] and [5.4] β_2 is the same as the β in [5.2] and $\beta_1 = \beta_2 \log \rho$ (see [5.2]). So, the estimate of
749 the relative potency, ρ , is

750

$$\hat{\rho} = \exp\left(\frac{\hat{\beta}_1}{\hat{\beta}_2}\right)$$

751

752 Commonly available statistical software and spreadsheets provide routines for least squares. Not all
753 software can provide weighting or mixed-model analyses.

754

755 **5.3.4 Nonlinear models.** Nonlinear dose–response models are typically S-shaped functions. They
756 occur when the range of doses is wide enough so that the lower doses have little or no response,
757 followed by a quickly rising response, and then reach an upper limit where the response reaches a

758 plateau. The most common of these are the four- and five-parameter logistic functions as given
 759 below.

760

761 Let y denote the observed response and z the concentration. One form of the four-parameter
 762 logistic model is

763
$$y = D + \frac{A - D}{1 + \left(\frac{z}{C}\right)^B} + e \quad [5.5]$$

764 One alternative, but equivalent, form is

765
$$y = a_0 + \frac{d}{1 + \exp[M(\ln z - b)]} + e \quad [5.6]$$

766

767 The two forms correspond as follows:

768 Lower asymptote: $D = a_0$

769 Upper asymptote: $A = a_0 + d$

770 Steepness: $B = M$ (related to the slope of the curve at the EC50)

771 Effective concentration 50% (EC50): $C = \exp(b)$ (may also be termed ED50).

772 The “ln” denotes natural log. Any convenient base for logarithms is suitable; it is often convenient
 773 to work in log base 2, particularly when doses are twofold apart; see Example 7.1.

774

775 The four-parameter logistic has the property of mirror-image symmetry around EC50 when plotted
 776 against log dose, and, in particular, the rates of approach to the upper and lower asymptotes are the
 777 same. See Figure 5.2. The five-parameter logistic is a model that relaxes this symmetry:

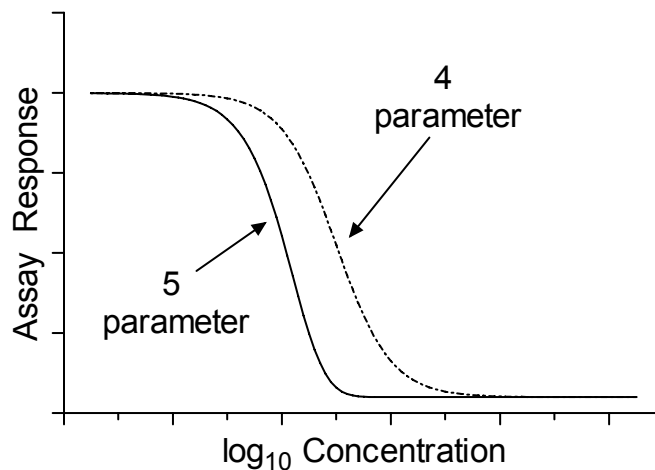
778
$$y = D + \frac{A - D}{\left[1 + \left(\frac{z}{C}\right)^B\right]^g} + e \quad [5.7]$$

779 Note that when $g \neq 1$, C can no longer be interpreted as EC50.

780

781

Figure 5.2. Examples of 4- and 5-parameter logistic curves



782

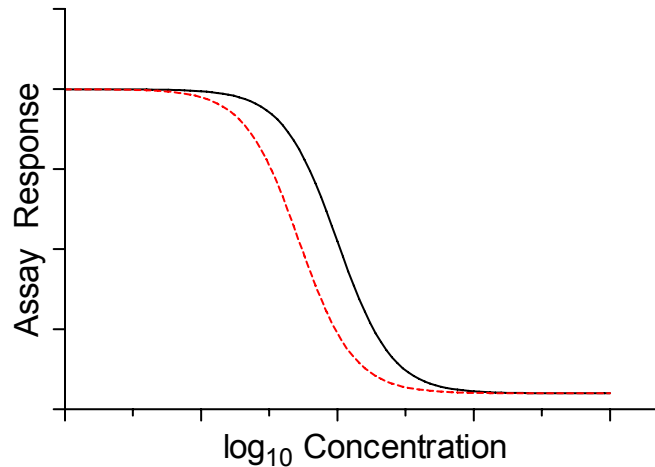
783

784 In many assays the analyst has a number of strategic choices. For example, the responses could be
 785 modeled using a transformed response fit to a four-parameter logistic curve, or the responses could

786 be weighted and fit to a five-parameter logistic curve. Several complex trade-offs are associated with
 787 these choices. It is often important to include terms in the model (often random effects) to address
 788 variation in the responses (or parameters of the response) associated with blocks or experimental
 789 units in the design of the assay. For simple assays where observations are independent, these
 790 strategic choices are fairly straightforward. For assays performed with grouped dilutions (i.e., via
 791 multichannel pipettes), with serial dilutions, or with designs that include blocks (i.e., multiple
 792 plates/assay) it is important to employ appropriate statistical methodology.

793
 794 **5.3.4.1 Parallel curve dose–response models.** The concept of parallelism is not specific to linear
 795 models. For nonlinear curves, parallel or similar means the dose–response curves are
 796 superimposable following a horizontal displacement of one of the curves, as shown in Figure 5.3 for
 797 four-parameter logistic curves.

798
 799 Figure 5.3. Example of parallel curves from a nonlinear model



800
 801
 802 The equations corresponding to the figure (with error terms added to the equations) are

$$y_R = D + \frac{A - D}{1 + \left(\frac{z}{C}\right)^B} + e$$

803 [5.8]

$$y_T = D + \frac{A - D}{1 + \left(\frac{\rho z}{C}\right)^B} + e$$

804 or

$$y_R = D + \frac{A - D}{1 + \exp\left[M(\ln z - b)\right]} + e$$

805 [5.9]

$$y_T = D + \frac{A - D}{1 + \exp\left[M(\ln z - b + \ln \rho)\right]} + e$$

806 In ρ is the natural log of the relative potency and the horizontal distance between the two curves.
 807 Because the EC50 of the standard is $\exp(b)$ and that of the Test is $\exp(b - \ln \rho)$, the relative potency is
 808 the ratio of EC50's (standard over Test) when the parallel curve model holds.

809

810 **5.3.4.2 Estimation of Parallel Curve Models.** Estimation of a nonlinear, parallel curve model is
 811 similar to that for parallel-line models. The method is still least squares, possibly after transformation
 812 and possibly with weighting. For the four-parameter logistic model, the parameter estimates are
 813 found by minimizing:

$$814 \quad \sum \left(y - \hat{D} - \frac{\hat{A} - \hat{D}}{1 + \exp \left[\hat{M} (\ln z - \hat{b} + \hat{r}T) \right]} \right)^2 \quad [5.10]$$

815 without weighting, or

$$816 \quad \sum \frac{1}{w} \left(y - \hat{D} - \frac{\hat{A} - \hat{D}}{1 + \exp \left[\hat{M} (\ln z - \hat{b} + \hat{r}T) \right]} \right)^2 \quad [5.11]$$

817 with weighting. In either case, the estimate of r is the estimate of the natural log of the relative
 818 potency. For some software, it may be easier to work with $a_0 = A - D$.

819 The parameters of the logistic functions, both four- and five-parameter, cannot be found with
 820 ordinary least squares regression routines. Computer programs with nonlinear estimation techniques
 821 must be used.

822 **5.3.5 Slope ratio dose–response models.** If the concentration–response model [5.1] can be made
 823 linear in $x = z^\lambda$, where z is dose or concentration for $\lambda \neq 0$, the resulting equations are then:

$$824 \quad \begin{aligned} 825 \quad y_R &= \alpha + \beta z^\lambda + e = \alpha + \beta x + e \\ 826 \quad y_T &= \alpha + \beta (\rho z)^\lambda + e = \alpha + \beta \rho^\lambda x + e \end{aligned} \quad [5.12]$$

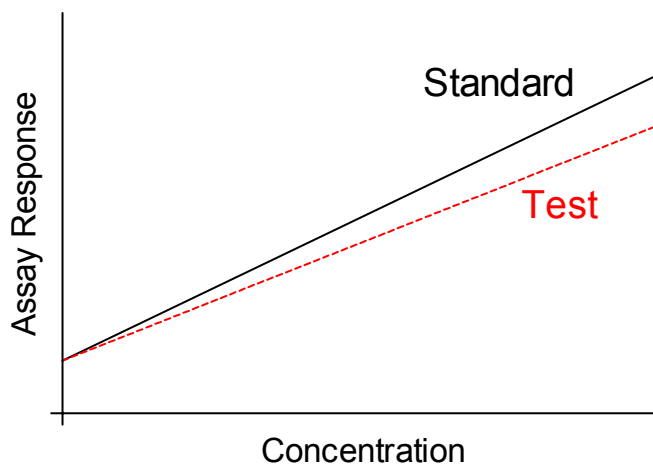
827 Thus, there are two identifying characteristics of a slope ratio dose–response model that can be seen
 828 only in the results of a ranging study:

- 831 1. The dose is not log transformed
- 832 2. The lines for different potencies from a ranging study all have the same intercept and
 833 have different slopes. Thus, a graph of the ranging study resembles a fan. Figure 5.4
 834 shows an example of a ranging study with a slope ratio dose–response model. The
 835 Test (solid line) has a relative potency of 2 relative to the Standard (dashed line).
 836 Note that the common intercept need not be the origin.

837

838

Figure 5.4. Example of slope ratio model



839

840

841 An assay with a slope-ratio dose–response model for measuring relative potency consists, at a
 842 minimum, of one Standard sample and one Test sample, each measured at three or more doses. The
 843 similarity test for this model is that the intercepts of the two lines are equivalent. (This is the analog
 844 of parallelism for slope-ratio models, and the considerations of Section 4.2 apply here as well.) A
 845 linear regression model using ordinary least squares regression can be fit to both lines
 846 simultaneously. The model consists of one common intercept, a slope for the Test sample results,
 847 and a slope for the Standard sample results. The relative potency is then found from the ratio of the
 848 slopes:

849

$$850 \quad (\text{Relative Potency})^\lambda = \frac{\text{Test sample slope}}{\text{Reference sample slope}} = \frac{\beta\rho^\lambda}{\beta} = \rho^\lambda \quad [5.13]$$

851

852 Let b_R represent the estimate of β , b_T represent the estimate of $\beta\rho^\lambda$, and R represent the estimate of
 853 the relative potency. Then

854

$$R = \left(\frac{b_T}{b_R} \right)^{\frac{1}{\lambda}}$$

855 Most often, $\lambda = 1$.

856

857 **5.3.5.1 Assumptions for and estimation of slope-ratio models.** The assumptions for the slope-
 858 ratio model are the same as for parallel-line models: The residual terms are independent, have
 859 constant variance, and may need to have a normal distribution. The method of estimation is also by
 860 the method of least squares. This may be implemented either with or without weighting, as
 861 demonstrated in equations [5.14] and [5.15], respectively.

$$862 \quad \sum (y - \hat{\alpha} - \hat{\beta}_R x(1-T) - \hat{\beta}_T xT)^2 \quad [5.14]$$

863

$$864 \quad \sum \frac{1}{w} \left(y - \hat{\alpha} - \hat{\beta}_R x(1-T) - \hat{\beta}_T xT \right)^2 \quad [5.15]$$

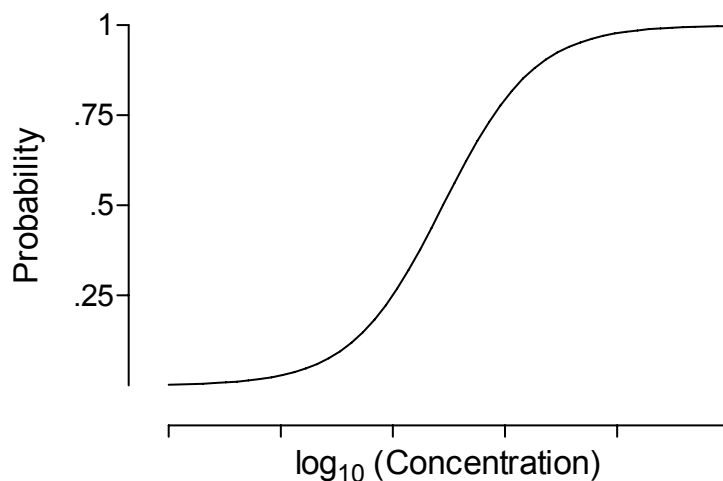
865 This is a linear regression with two independent variables, $x(1-T)$ and xT . $\hat{\beta}_T$ is then the estimated
 866 slope for the Test and $\hat{\beta}_R$ the estimated slope for the Standard, and the estimate of relative potency
 867 is $R = \hat{\beta}_T / \hat{\beta}_R$.

868 Because the slope-ratio model is a linear regression model, most statistical packages and
 869 spreadsheets can be used to obtain the relative potency estimate. In some assay systems, sometimes
 870 on an assay instance-specific basis it is appropriate to omit the zero dose and at times one or more
 871 of the high doses. A number of strategic and technical issues are involved in developing a good
 872 method for selecting the subset of doses for a slope-ratio assay.
 873

874 **5.4 Dichotomous (Quantal) Assays.** For quantal assays the assay measurement has a dichotomous
 875 or binary outcome. In animal assays, this may be that the animal is dead or alive or that a certain
 876 physiologic response is observed or not. For cellular assays, the quantal response may be whether
 877 there is a response in the cell or not. In cell-based viral titer or colony-forming assays, the quantal
 878 response may be a limit of integer response, such as an integer number of particles or colonies. If
 879 what can be readily determined is whether there are any particles—but not their actual number—
 880 then the assay can be analyzed as quantal. Note that if the reaction can be quantitated on a
 881 continuous scale, as with an optical density, then the assay is not quantal.
 882

883 **5.4.1 Models for quantal analyses.** The key to models for quantal responses is to work with the
 884 probability of a response (e.g., probability of death), in contrast to quantitative responses where the
 885 model is for the response itself. For each dose, d , a treated animal has a probability of responding to
 886 that dose, $P(d)$. Often the curve $P(d)$ can be approximated by a sigmoid when plotted against the
 887 logarithm of dose, as shown in Figure 5.5. This curve captures that the probability of responding
 888 increases with dose. The dose that corresponds to a probability of 0.5 is the *ED50* or the dose at
 889 which 50% of animals are expected to respond.
 890

891 Figure 5.5. Example of sigmoid for $P(d)$
 892



893

894
 895 The sigmoid curve is usually modeled based on either the normal distribution or the logistic
 896 distribution. If the normal distribution is used, the resulting analysis is termed probit analysis; if the
 897 logistic is used, the analysis is termed logit or logistic analysis. The probit and logit models are
 898 practically indistinguishable; either is an acceptable choice. The choice may be based on the
 899 availability of software that meets the laboratory's analysis and reporting needs. Because software is
 900 more commonly available for logistic models (often under the term logistic regression), this
 901 discussion will focus on the use and interpretation of logit analysis. The same considerations
 902 discussed in this Section for logit analysis apply as well to probit analysis; see example 7.5.

903
 904 **5.4.1.1 Logit model.** The logit model for the probability of response, $P(d)$, can be expressed in two
 905 equivalent forms. For the sigmoid,

$$\begin{aligned}
 P(d) &= \frac{\exp[\beta_0 + \beta_1 \ln(d)]}{1 + \exp[\beta_0 + \beta_1 \ln(d)]} \\
 &= \frac{\exp[\beta_1 \ln(d / ED50)]}{1 + \exp[\beta_1 \ln(d / ED50)]}, \text{ where } \ln(ED50) = -\beta_0 / \beta_1 \\
 &= \frac{1}{1 + \exp[-\beta_1 \ln(d / ED50)]} \\
 &= \frac{1}{1 + (d/ED50)^{-\beta_1}}
 \end{aligned}
 \tag{5.16}$$

906
 907 An alternative form shows the relationship to linear models:

$$\text{logit transform of } P = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \ln(d)
 \tag{5.17}$$

908
 909 The linear form is a useful reminder that many of the considerations discussed in 5.3.3 for linear
 910 models apply to quantal models as well, in particular linearity and parallelism.

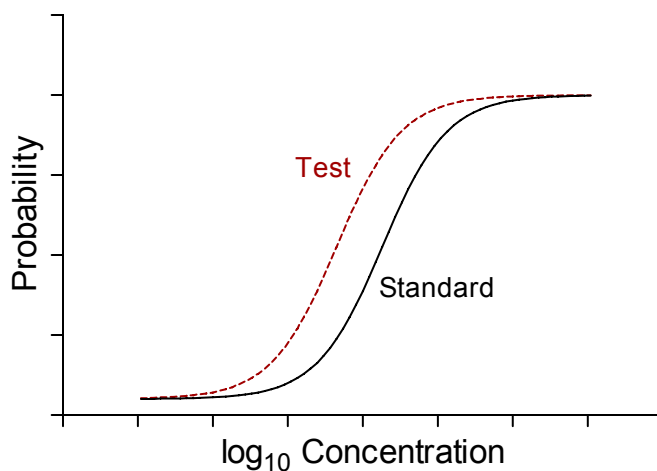
911
 912 For a logit analysis with Standard and Test preparations, let T be a variable that takes the value 1 for
 913 animals receiving the Test preparation and 0 for animals receiving the Standard. Assuming
 914 parallelism of the Test and Standard preparations, the logit model for estimating relative potency is
 915 then:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \ln(d) + \beta_2 T
 \tag{5.18}$$

916
 917
 918
 919 The natural log of the relative potency of the Test compared to the Standard preparation is then
 920 β_2 / β_1 . The two curves in Figure 5.6 show parallel Standard and Test sigmoids. (If the
 921 corresponding linear forms were shown, they would be two parallel straight lines.) Here, the scale
 922 was chosen so the relative potency of the Test to the Standard was 1.5. The log of the relative
 923 potency is the horizontal distance between the two curves, in the same way as for the linear and
 924 four-parameter logistic models given for quantitative responses (Sections 5.3.2.3 and 5.3.3.1).
 925
 926

927

Figure 5.6. Example of Parallel Sigmoid Curves

928
929

930 **5.4.2 Estimating the model parameters and relative potency.** There are two methods for
 931 estimating the parameters of logit and probit models: maximum likelihood and weighted least
 932 squares. The least squares approach is the one most commonly used for probit analysis. Software for
 933 logit analysis that is intended for bioassay application is also likely to use weighted least squares.
 934 More general logistic software will most likely use maximum likelihood. The difference is not
 935 practically important, and the laboratory can accept the choice made by its software. The following
 936 assumes a general logistic regression software program. Specialized software should be similar.

937
 938 Considering the form of Equation (5.17), one observes a resemblance to linear regression. This is a
 939 useful perspective if logistic regression software is used. There are two independent variables, $\ln(d)$
 940 and T . For each animal, there is a yes/no dependent variable, often coded as 1 for yes or *response* and
 941 0 for *no* or *no response*. Although bioassays are often designed with equal numbers of animals per
 942 dose, that is not a requirement of analysis. Utilizing the parameters estimated by software, which
 943 include β_0 , β_1 , and β_2 and their Standard errors, one obtains the estimate of the natural log of the
 944 relative potency:

$$945 \quad \text{Estimate of natural log of relative potency} = \frac{\hat{\beta}_2}{\hat{\beta}_1} \quad [5.19]$$

946 where the caret denotes estimate. Substituting the parameter estimates and their Standard errors into
 947 Fieller's Theorem (see Section 4.10), one can calculate the confidence interval $[L, U]$ for the log of
 948 relative potency. The confidence interval for the relative potency is then $[\exp(L), \exp(U)]$.

949
 950 **5.4.3 Assumptions.** In most cases, quantal analyses assume a Standard binomial model with
 951 independence of results from animal to animal. The binomial is a common choice of distribution for
 952 dichotomous data in contrast to the choice of normal distribution for continuous data. The key
 953 assumptions of the binomial are that at a given dose each animal treated at that dose has the same
 954 probability of responding, and the results of any animal are not correlated with those of any other
 955 animal. There are many ways in which this basic set of assumptions could be violated. Foremost
 956 among these would be the presence of litter effects, where animals from the same litter will tend to
 957 respond more alike than will animals from different litters. Cage effects, in which the environmental

958 conditions or care rendered to any specific cage makes the animals from that cage more or less likely
 959 to respond to experimental treatment, violates the same equal-probability assumption. These
 960 assumption violations and others like them (that could be a deliberate design choice) do not
 961 preclude the use of logit or probit models; however, they are indications that a more complex
 962 approach to analysis than that presented here is required.

964 **5.4.3.1 Assessing assumptions.** To assess parallelism, equation [5.17] may be modified as follows:

$$965 \quad \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \ln(d) + \beta_2 T + \beta_3 T * \ln(d) \quad [5.20]$$

966 Here, β_3 is the difference of slopes between Test and Standard and should be sufficiently small. (The
 967 $T*\ln(d)$ term is known as an *interaction term* in statistical terminology.) The measure of nonparallelism
 968 may also be expressed in terms of the ratio of slopes, $(\beta_1 + \beta_3)/\beta_1$. See Section 4.2 for a general
 969 discussion of assessing parallelism.

970
 971 To assess linearity, it is good practice to start with a graphical examination. In accordance with
 972 [5.17], this would be a plot of $\ln[(y+.5)/(n-y+.5)]$ against $\ln(\text{dose})$, where y is the total number of
 973 responses at the dose and n is the number of animals at that dose. The lines for Test and Standard
 974 should be parallel straight lines as for the linear model in quantitative assays.

975
 976 Formal tests for linearity are based on testing the null hypothesis of linearity against the alternative
 977 of nonlinearity. As discussed during the assessment of parallelism, an alternative method that
 978 supports a conclusion of *sufficiently linear* is favored.

979
 980 Linearity of the $\ln(\text{dose})$ term in [5.17] is not a necessity. This contrasts to parallelism, which is a key
 981 assumption of a relative potency assay. The dose–response relationship can be examined by plotting
 982 $\ln((\% \text{response})/(100 - \% \text{response}))$ versus $\ln(\text{dose})$. If the relationship is monotonic but does not
 983 appear to be linear, then the model in [5.17] can be extended with other terms. For example, a
 984 quadratic term in $\ln(\text{dose})$ could be added: $[\ln(\text{dose})]^2$. If dose needs to be transformed to something
 985 other than log dose, then the quantal model analogue of slope ratio assays is present. The latter are
 986 possible but are sufficiently unusual that they will not be discussed here in any greater depth.

987
 988 **5.4.4 Outliers.** The concept of outliers is different for quantal assays than for quantitative assays—
 989 because the assay response is either yes or no, the value cannot be unusual. What may appear to fall
 990 into the outlier category is a single response at a low dose or a single no-response at a high dose.
 991 Assuming that there has been no cause found (e.g., failure to properly administer the drug to the
 992 animal), there is no statistical basis for distinguishing an outlier from a rare event.

993
 994 **5.4.5 Alternative methods.** Alternatives to the simple quantal analyses outlined here may be
 995 acceptable, depending on the nature of the analytical challenge. One such challenge is a lack of
 996 independence among experimental units, as may be seen in litter effects in animal assays.

997
 998 Three approaches that may be employed are Generalized Estimating Equations (GEE), Generalized
 999 linear models, and generalized linear mixed-effects models. A GEE analysis will yield standard errors
 1000 and confidence intervals whose validity does not depend on the satisfaction of the independence
 1001 assumption.

1002

6.0 COMBINING RESULTS FROM MULTIPLE ASSAYS

6.1 Preliminary Considerations. In order to mitigate the effects of variability, replication of independent bioassays and combination of their results to obtain a single reportable value is often necessary. Analysts should evaluate whether it is appropriate to combine the results of such assays and, if so, in what way to proceed.

There are two primary questions to address when considering whether to combine results from multiple assays:

Are the assays mutually independent?

A set of assays may be regarded as mutually independent when the execution of one does not affect the possible outcomes of any of the others. This implies that the random errors in all essential factors influencing the result (for example, dilutions of the standard and of the preparation to be examined or the sensitivity of the biological indicator) in one assay must be independent of the corresponding random errors in the other one. Assays on successive days using the original and retained dilutions of the standard, therefore, are not independent assays.

It is not a requirement that assays be independent in order for analysts to combine results. However, methods for independent assays are much simpler. Also, combining dependent assay results may require assumptions about the form of the correlation between assay results that may not be verifiable. Statistical methods are available for dependent assays, but they are not presented here.

Are the results of the assays homogeneous?

Homogeneous results differ only due to random within-assay errors. Any contribution from factors associated with intermediate precision precludes homogeneity of results. Intermediate precision factors are those that vary between assays within a laboratory and can include analyst, equipment, and environmental conditions. There are statistical tests for heterogeneity, but lack of statistically significant heterogeneity is not properly taken as assurance of homogeneity. Homogeneity should be documented by examining data from multiple assays.

Additionally, before results from assays can be combined, analysts should consider the scale on which that combination is to be made. In general, the combination should be done on the scale for which the parameter estimates are approximately normally distributed. Thus, for relative potencies based on a parallel-line or parallel-curve method, the relative potencies are combined in the logarithm scale.

6.2 Methods for Combining Assays. There are several methods for combining the results of independent assays, the choice among which depends on judgments regarding assay result homogeneity. A simple approximate method is described below and is recommended. A second procedure is provided and may be useful if necessary conditions are fulfilled and documented. A third alternative, analyzing all assays together using a linear or nonlinear mixed-effects model, is not discussed here.

1051
 1052 **6.2.1 Method 1—assay result homogeneity not required.** Following is a simple method that
 1053 assumes independence of assays but not homogeneity of results. It is further assumed that the
 1054 individual assay results are from a common normal distribution. This latter assumption requires that
 1055 all assays to be combined used the same design and laboratory procedures.

1056
 1057 Let R_i denote the result (on the appropriate scale; this will usually be the logarithm of the relative
 1058 potency) of the i^{th} assay of N assay results to be combined. To combine the N results, the mean,
 1059 standard deviation, and standard error are calculated in the usual way:

$$\text{Mean } \bar{R} = \sum_{i=1}^N R_i / N$$

1060
 Standard Deviation $S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (R_i - \bar{R})^2}$

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

1061
 1062 A $100(1 - \alpha)\%$ confidence interval is then found as

$$\bar{R} \pm t_{N-1, \alpha/2} SE$$

1063
 1064 where $t_{N-1, \alpha/2}$ is the upper $\alpha/2$ percentage point of a t -distribution with $N-1$ degrees of freedom.

1065 The quantity $t_{N-1, \alpha/2} SE$ is the expanded uncertainty of \bar{R} . The number, N , of assays to be
 1066 combined is usually small, and hence the value of t is usually large.

1067
 1068 If the result is combined in the logarithm scale, results can be reported in the untransformed scale as
 1069 a confidence interval for the geometric mean,

1070

$$\exp(\bar{R} - t_{N-1, \alpha/2} SE), \exp(\bar{R} + t_{N-1, \alpha/2} SE).$$

1071
 1072 **6.2.2 Method 2—assay result homogeneity required.** This method can be used provided the
 1073 following conditions are fulfilled:

- 1074 (1) The individual potency estimates form a homogeneous set with regard to the potency
 1075 being estimated. Note that this means documenting that there are no contributions to
 1076 between-assay variability from intermediate precision factors. Standard statistical testing
 1077 of a null hypothesis of no contribution is not appropriate for demonstrating this
 1078 condition; the contribution should be demonstrated to be equivalent to zero.
- 1079 (2) Potency estimates are derived from independent assays.
- 1080 (3) The number of degrees of freedom of the individual residual errors is not smaller than 6
 1081 and preferably is larger than 15.

1082
 1083 When these conditions are not fulfilled, this method cannot be applied and Method 1 or some other
 1084 method should be used. Further note that although Method 2 often results in narrower confidence
 1085 intervals than Method 1, this is not sufficient justification for using Method 2 absent satisfaction of
 1086 the conditions listed above.

1087
 1088 **6.2.2.1 Calculation of weighting coefficients.** It is assumed that the results of each of the N
 1089 assays have been analyzed to give N values of \log *potency* with associated confidence limits. For each

1090 assay, i , the logarithmic confidence interval for the log potency and a value L_i are obtained by
 1091 subtracting the lower limit from the upper. A weight W_i for each value of the log relative potency,
 1092 R_i , is calculated as follows, where t_i has the same value as that used in the calculation of confidence
 1093 limits in the i^{th} assay.

1094
$$W_i = \frac{4t_i^2}{L_i^2}$$

1095
 1096 **6.2.2.2 Calculation of the weighted mean and confidence limits.** The products $W_i R_i$ are
 1097 formed for each assay, and their sum is divided by the total weight for all assays to give the
 1098 weighted mean log relative potency and its standard error as follows:

1099
$$\text{Mean } \bar{R} = \sum_{i=1}^N W_i R_i / \sum_{i=1}^N W_i$$

Standard Error $SE = 1 / \sqrt{\sum_{i=1}^N W_i}$

1100 A 100(1- α)% confidence interval in the log scale is then found as

1101
$$\bar{R} \pm t_{k,\alpha/2} SE$$

1102 where $t_{k,\alpha/2}$ is the upper $\alpha/2$ percentage point of a t -distribution with degrees of freedom, k , equal to
 1103 the sum of the number of degrees of freedom for the error mean squares in the individual assays.
 1104 This confidence interval can then be transformed back to the original scale as for Method 1.

1105
 1106 **7.0 EXAMPLES**

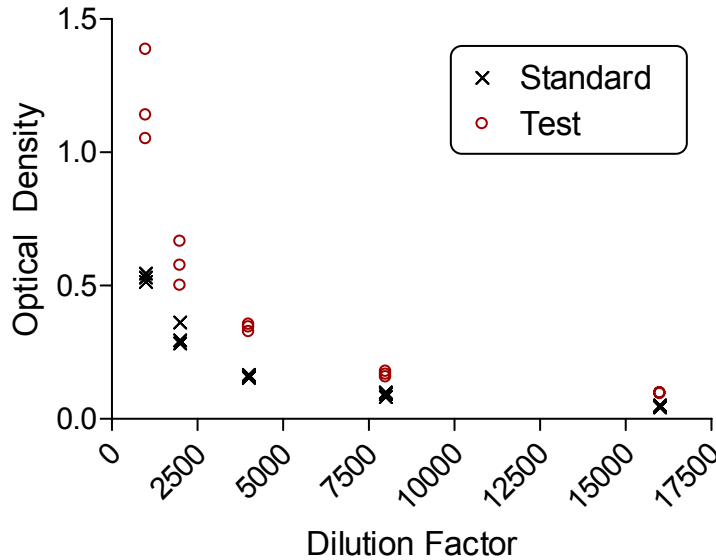
1107
 1108 **7.1 Example of Parallel-line Analysis.** An in vitro assay was conducted to compare a new hepatitis
 1109 B vaccine against a Standard vaccine. Three independent two-fold dilution series of 5 dilutions were
 1110 prepared from each of the vaccines. After additional assay steps, optical densities were measured.
 1111 These are listed in Table 7.1.1.

1112
 1113
 1114 **Table 7.1.1. — Optical Densities (OD)**

Dilution	Standard			Test		
	1:16 000	0.043	0.045	0.051	0.097	0.097
1:8000	0.093	0.099	0.082	0.167	0.157	0.178
1:4000	0.159	0.154	0.166	0.327	0.355	0.345
1:2000	0.283	0.295	0.362	0.501	0.665	0.576
1:1000	0.514	0.531	0.545	1.140	1.386	1.051

1116
1117

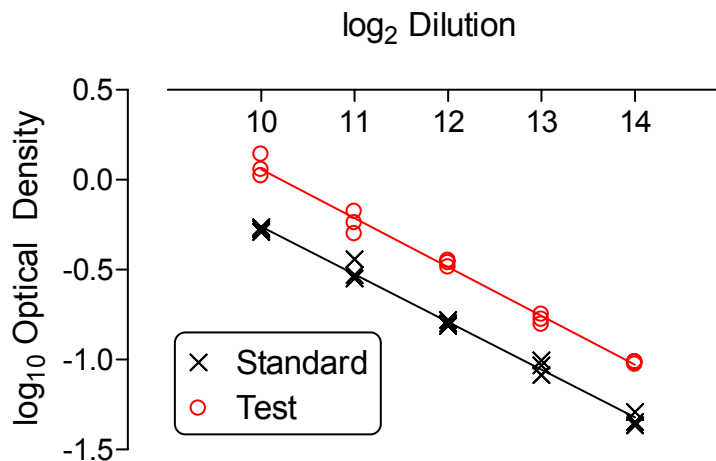
Figure 7.1.1. Plot of OD vs. 1/Dilution Factor



1118
1119
1120
1121
1122
1123
1124
1125
1126

Figure 7.1.1 illustrates two problems. First, the relationship between OD and dilution factor is not linear. Second, observation suggests that at least for the Test data, the variability increases with OD and that a weighted analysis may be needed. Figure 7.1.2 shows the same data but now with a base 2 log scale for dilution. Because the dilutions are by successive factors of 2, assigning successive integer values to the dilutions is the same as using a \log_2 scale and gives us the convenience of integers for the logs of dilution factors.

Figure 7.1.2. Plot of \log_{10} OD vs. \log_2 Dilution



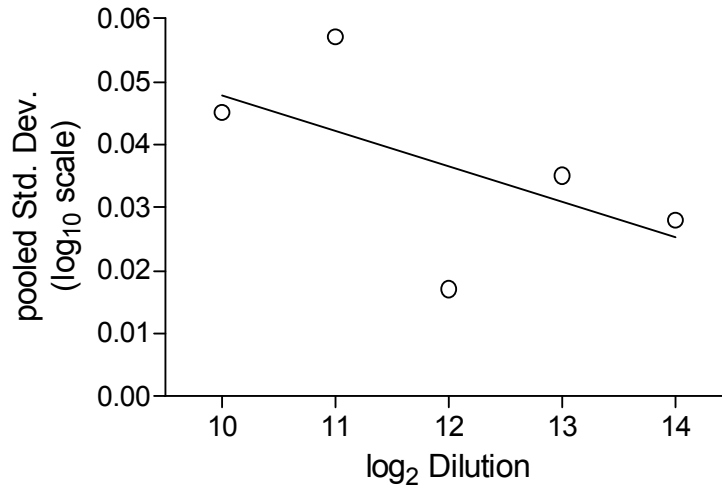
1127
1128
1129
1130
1131
1132
1133

Following transformation, the relationship is now reasonably linear; also, the variability now appears more constant across the plot. Insofar as linearity, the primary motivator of transformation, has been achieved, no further transform will be used to address inconstant variance. A further inspection of variability is obtained by plotting standard deviations (standard deviations where Test and standard variances are pooled) at each successive dilution against dilution. Figure 7.1.3 depicts

1134 this association with a trend line (regression) added. For the purpose of this exercise, this apparent
 1135 trend is deemed sufficient cause to require compensation by weighting.

1136
 1137
 1138

Figure 7.1.3. Standard Deviation vs. Dilution



1139
 1140

1141 How is a weighting factor determined? First, it is inappropriate to use the standard deviations
 1142 plotted in Figure 7.1.3 as the basis of weighting. Because they are based on scant data—two sets of
 1143 triplicates—they are insufficient for an accurate determination. If historical data are available, they
 1144 could be used to estimate the variance at each dilution and could be used for the weighting. Another
 1145 option, whether historical data are available or not, is to model the relationship between the variance
 1146 and dilution and use that information to set a weighting factor. Here, the standard deviation is
 1147 roughly linearly associated with dilution and, after some trial and error, roughly proportional to (9-
 1148 fold dilution as coded in Figure 7.1.3). Again, the variance at each dilution is not required, only a
 1149 relationship based on the entire model (regression) association variance with dilution across the
 1150 entire range of dilutions. Thus, the variance is modeled as proportional to (9-fold dilution)² and
 1151 responses are weighted inversely proportional to that value.

1152

1153 A weighted regression to the log transformed ODs is now fit, with two slopes, an intercept, and a
 1154 difference of intercepts. Table 7.1.2 shows the data in a convenient spreadsheet format for this
 1155 analysis. The variable Test is 1 for data from the Test vaccine and 0 for data from the Standard. This
 1156 arrangement is for spreadsheets and other programs that do not have specific functions for weighted
 1157 regression but can determine a regression with the intercept forced to be 0. The prime variables are
 1158 original variables multiplied by the square root of the weight. *TD* indicates Test_Dilution, the
 1159 product of the Test and log dilution variables; *D* stands for Dilution. If this analysis were performed
 1160 without weighting, the first four columns would be used and a non-zero intercept would be sought.
 1161 With weighting, the last five columns are used and an intercept of zero is sought.

1162

Table 7.1.2. Data in Spreadsheet Format

LOG ₁₀ (OD)	Test_ Dilution	log ₂ dilution	Test	Weight	TD'	D'	TEST'	SQRT(Weight)	LOG ₁₀ (OD)'
-1.3665	0	5	0	0.0625	0	1.25	0	0.25	-0.3416
-1.3468	0	5	0	0.0625	0	1.25	0	0.25	-0.3367
-1.2924	0	5	0	0.0625	0	1.25	0	0.25	-0.3231
-1.0315	0	4	0	0.0400	0	0.8	0	0.2	-0.2063
-1.0044	0	4	0	0.0400	0	0.8	0	0.2	-0.2009
-1.0862	0	4	0	0.0400	0	0.8	0	0.2	-0.2172
-0.7986	0	3	0	0.0278	0	0.5	0	0.1667	-0.1331
-0.8125	0	3	0	0.0278	0	0.5	0	0.1667	-0.1354
-0.7799	0	3	0	0.0278	0	0.5	0	0.1667	-0.1300
-0.5482	0	2	0	0.0204	0	0.2857	0	0.1429	-0.0783
-0.5302	0	2	0	0.0204	0	0.2857	0	0.1429	-0.0757
-0.4413	0	2	0	0.0204	0	0.2857	0	0.1429	-0.0630
-0.2890	0	1	0	0.0156	0	0.125	0	0.125	-0.0361
-0.2749	0	1	0	0.0156	0	0.125	0	0.125	-0.0344
-0.2636	0	1	0	0.0156	0	0.125	0	0.125	-0.0330
-1.0132	5	5	1	0.0625	1.25	1.25	0.25	0.25	-0.2533
-1.0132	5	5	1	0.0625	1.25	1.25	0.25	0.25	-0.2533
-1.0269	5	5	1	0.0625	1.25	1.25	0.25	0.25	-0.2567
-0.7773	4	4	1	0.0400	0.8	0.8	0.2	0.2	-0.1555
-0.8041	4	4	1	0.0400	0.8	0.8	0.2	0.2	-0.1608
-0.7496	4	4	1	0.0400	0.8	0.8	0.2	0.2	-0.1499
-0.4855	3	3	1	0.0278	0.5	0.5	0.1667	0.1667	-0.0809
-0.4498	3	3	1	0.0278	0.5	0.5	0.1667	0.1667	-0.0750
-0.4622	3	3	1	0.0278	0.5	0.5	0.1667	0.1667	-0.0770
-0.3002	2	2	1	0.0204	0.2857	0.2857	0.1429	0.1429	-0.0429
-0.1772	2	2	1	0.0204	0.2857	0.2857	0.1429	0.1429	-0.0253
-0.2396	2	2	1	0.0204	0.2857	0.2857	0.1429	0.1429	-0.0342
0.0569	1	1	1	0.0156	0.125	0.125	0.125	0.125	0.0071
0.1418	1	1	1	0.0156	0.125	0.125	0.125	0.125	0.0177
0.0216	1	1	1	0.0156	0.125	0.125	0.125	0.125	0.0027

1163

1164 The results are shown in Table 7.1.3.

1165

1166

Table 7.1.3. Regression Results with Unequal Slopes

	Coefficients	Standard Errors	<i>t</i> Statistics	<i>p</i> -values	Lower 95% CI ^a	Upper 95% CI ^a
Intercept	0	N/A ^b	N/A	N/A	N/A	N/A
TD'	-0.0019	0.00967	-0.198	0.844	-0.0218	0.0180
D'	-0.2681	0.00684	-39.21	<< 0.001	-0.28213	-0.2540
TEST'	0.3107	0.0379	8.208	<< 0.001	0.232917	0.3886
SQRT(weight)	0.0137	0.0268	0.510	0.614	-0.04137	0.0687

1167 ^aCI = confidence interval.1168 ^bN/A = not applicable.

1169

1170 With reference to the data in the format of Table 7.1.2, the coefficient for the variable TD' is the
 1171 difference of the slopes, the coefficient for $SQRT(\text{weight})$ is actually the intercept and is that for the
 1172 Standard, and the coefficient for $Test'$ is the difference of the intercepts. The lines in Figure 7.1.2
 1173 correspond to this analysis. The difference of slopes (-0.0019) is small and the 95% confidence
 1174 interval for the difference of slopes, (-0.0218, 0.0180), covers an interval that is a small fraction of
 1175 the Standard slope (-0.268). Therefore, a conclusion that the slopes are sufficiently similar to
 1176 proceed to an analysis with a common slope is warranted. Although it is also the case that the p -
 1177 value for the difference of slopes (0.844) indicates no statistically significant difference of slopes,
 1178 that is not relevant to the conclusion. Table 7.1.4 shows some of the results after refitting the
 1179 regression with a common slope. This uses the last four columns of Table 7.1.2.

1180
 1181 **Table 7.1.4. A Portion of the Results from Analysis Assuming Equal Slopes**

	Coefficients	Standard Errors
Intercept	0	N/A
D'	-0.2690	0.0047
TEST'	0.3037	0.0127
SQRT(weight)	0.0172	0.0196

1182
 1183 The estimate of the logarithm of relative potency is $-0.3037/(-0.2690) = 1.129$. The extra minus sign
 1184 is here because the dilution coding is reversed from concentration order; see the Figures. Taking
 1185 base 2 antilogs (because the base 2 log scale is used for the dilution variable), the estimate of relative
 1186 potency is 2.187. Fieller's Theorem is used to obtain a confidence interval. This requires the
 1187 covariance between the estimate of the slope and the estimate of the difference of intercepts (Test
 1188 row of Table 7.1.4). Zero is substituted for the vanishingly small covariance value of 4×10^{-20} , and the
 1189 simpler of the Fieller's formulas is used. Using the notation of Section 4.10:

1190 $\hat{R} = 1.129$

1191 $\hat{a} = -0.3037, SE_a = 0.0127$

1192 $\hat{b} = -0.2690, SE_b = 0.0047$

1193 $t = 2.052$, the upper 2.5% point of a t -distribution with 27 ($=30 - 3$) degrees of freedom (for a 95%
 1194 confidence interval)

1195 $g = (2.052 * 0.0047 / (-0.2690))^2 = 0.0013$.

1196

1197 The 95% confidence interval for the logarithm of relative potency is then

1198
$$\frac{1.129 \pm \frac{2.052}{-0.269} \sqrt{(1 - .0013)(0.0127)^2 + (1.129)^2 (0.0047)^2}}{1 - .0013} = (1.025, 1.236)$$

1199

1200 Taking base 2 antilogs, the 95% confidence interval for the relative potency is (2.04, 2.36).

1201

1202 In sum, a relative potency of 2.19 with a 95% confidence interval of (2.04, 2.36) has been
 1203 determined. This can be converted to a measure of specific activity using as a factor the Standard's
 1204 activity assignment of 20.0 μg protein/mL. This yields a specific activity of 43.7 for the Test
 1205 preparation with 95% confidence interval (CI) (40.8, 47.2).

1206

1207 Note that the analyses presented here assume independence. If serial dilutions had been used in
 1208 sample preparation, lack of that source of bias or variance renders these calculations suspect.

1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220

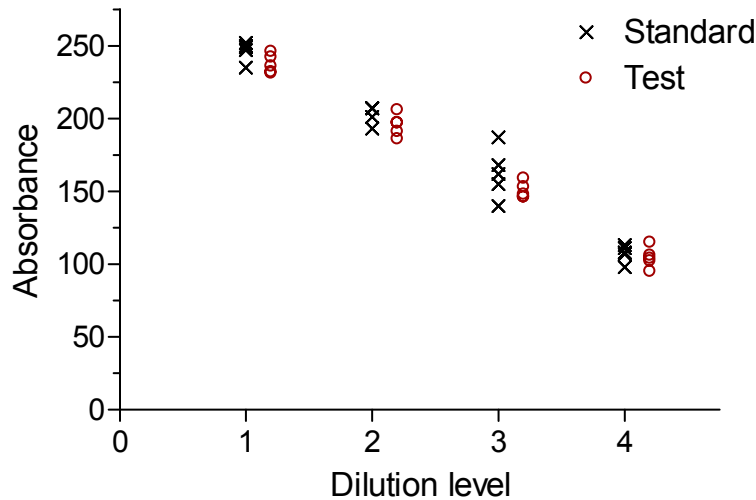
7.2 Example of Analysis of a Randomized Complete Block Design (RCBD). This assay is designed to assign a potency in international units (IU) per vial. The Standard has an assigned potency of 670 IU/mg. The Test preparation has an assumed potency of 20,000 IU/vial. On the basis of this information the stock solutions are prepared as follows: 16.7 mg of the Standard is dissolved in 25 mL solvent, and the contents of one vial of the Test preparation are dissolved in 40 mL solvent. The final solutions are prepared by first diluting to 1/40 and further using a dilution ratio of 1.5. The tubes are placed in a water bath in a randomized block arrangement see; <1032>. Blocking is used here to control potential variability associated with location in the bath. The responses are shown in Table 7.2.1 and plotted in Figure 7.2.1.

Table 7.2.1. Data for Randomized Block Example (absorbances x 1000)

Block	Standard <i>S</i>				Test <i>T</i>				Mean
	<i>S</i> ₁	<i>S</i> ₂	<i>S</i> ₃	<i>S</i> ₄	<i>T</i> ₁	<i>T</i> ₂	<i>T</i> ₃	<i>T</i> ₄	
1	252	207	168	113	242	206	146	115	181.1
2	249	201	187	107	236	197	153	102	179.0
3	247	193	162	111	246	197	148	104	176.0
4	250	207	155	108	231	191	159	106	175.9
5	235	207	140	98	232	186	146	95	167.4

1221
1222

Figure 7.2.1



1223
1224
1225
1226
1227
1228
1229
1230
1231

We see from the Figure that the linearity assumption is reasonable. We also see similar spread of the data at each dilution level, suggesting there is no need for a weighted analysis. As a reminder, normally the weighting would have been determined at or prior to validation and used in the analysis.

For analysis, we have two choices, namely whether or not to take the blocking into account. We could ignore the blocking and proceed as in Example 7.1. This would yield a valid estimate of relative potency but one with less precision than is possible. In the rows of Table 7.2.1, the mean

1232 absorbance decreases from block 1 to block 5, suggesting that there is a difference between blocks.
 1233 The purpose of the randomized block design is to increase precision by removing a source of
 1234 variability associated with the blocks and to provide a smaller error term and narrower confidence
 1235 intervals. Figure 7.2.2 shows the data in one convenient form for spreadsheets.
 1236
 1237

Table 7.2.2. Data in One Spreadsheet Format

Absorbances	Test_Dilution	Dilution	Test	block1	block2	block3	block4
252	0	1	0	1	0	0	0
249	0	1	0	0	1	0	0
247	0	1	0	0	0	1	0
250	0	1	0	0	0	0	1
235	0	1	0	0	0	0	0
207	0	2	0	1	0	0	0
201	0	2	0	0	1	0	0
193	0	2	0	0	0	1	0
207	0	2	0	0	0	0	1
207	0	2	0	0	0	0	0
168	0	3	0	1	0	0	0
187	0	3	0	0	1	0	0
162	0	3	0	0	0	1	0
155	0	3	0	0	0	0	1
140	0	3	0	0	0	0	0
113	0	4	0	1	0	0	0
107	0	4	0	0	1	0	0
111	0	4	0	0	0	1	0
108	0	4	0	0	0	0	1
98	0	4	0	0	0	0	0
242	1	1	1	1	0	0	0
236	1	1	1	0	1	0	0
246	1	1	1	0	0	1	0
231	1	1	1	0	0	0	1
232	1	1	1	0	0	0	0
206	2	2	1	1	0	0	0
197	2	2	1	0	1	0	0
197	2	2	1	0	0	1	0
191	2	2	1	0	0	0	1
186	2	2	1	0	0	0	0
146	3	3	1	1	0	0	0
153	3	3	1	0	1	0	0
148	3	3	1	0	0	1	0
159	3	3	1	0	0	0	1
146	3	3	1	0	0	0	0
115	4	4	1	1	0	0	0
102	4	4	1	0	1	0	0
104	4	4	1	0	0	1	0
106	4	4	1	0	0	0	1
95	4	4	1	0	0	0	0

1238

1239 Table 7.2.3 shows the results from the regression analysis with unequal slopes. One also sees the
 1240 block trend. The coefficient for Block 5 is identically 0, a consequence of the coding for block of
 1241 Table 7.2.2.

1242 **Table 7.2.3. Regression Results, Unequal Slopes**

	Coefficients	Standard Errors	<i>t</i> Statistics	<i>p</i>- values	Lower 95% CI	Upper 95% CI
Intercept	285.900	4.702	60.802	<<0.001	276.322	295.478
Test_Dilution	1.420	2.103	0.675	0.504	-2.863	5.703
Dilution	-45.820	1.487	-30.815	<<0.001	-48.849	-42.791
Test	-11.500	5.759	-1.997	0.054	-23.230	0.230
block1	13.750	3.717	3.699	0.001	6.178	21.322
block2	11.625	3.717	3.127	0.004	4.053	19.197
block3	8.625	3.717	2.320	0.027	1.053	16.197
block4	8.500	3.717	2.287	0.029	0.928	16.072

1243
 1244 The difference of slopes (1.420) is small, but there could be a question regarding the confidence
 1245 interval. The upper confidence bound (5.703) is more than 10% of the Standard slope (-45.820). For
 1246 most purposes, a difference of less than 20% or so is unlikely to be important, so we take this as
 1247 indicating sufficient parallelism that the analysis can continue with equal slopes. (The choice of 20%
 1248 is for this example and is not intended as generally applicable.) The equal-slopes analysis is based on
 1249 (5.3) with the addition of terms for the blocking. Those results are given in Table 7.2.4.

1250
 1251 **Table 7.2.4. Regression Results, Equal Slopes**

	Coefficients	Standard Errors	<i>t</i> Statistics	<i>p</i>- values	Lower 95% CIs	Upper 95% CIs
Intercept	284.125	3.867	73.483	<<0.001	276.259	291.991
Dilution	-45.110	1.043	-43.262	<<0.001	-47.231	-42.989
Test	-7.950	2.332	-3.410	0.002	-12.694	-3.206
block1	13.750	3.687	3.730	0.001	6.250	21.250
block2	11.625	3.687	3.153	0.003	4.125	19.125
block3	8.625	3.687	2.340	0.026	1.125	16.125
block4	8.500	3.687	2.306	0.028	1.000	16.000

1252
 1253 At this point analysis can proceed as in Example 7.1. The error mean square for the analysis of
 1254 Table 7.2.4 is 54.36; had blocks not been included as a factor in the analysis, this value would be
 1255 72.18.

1256
 1257 Using the notation of Section 4.10,

1258 $\hat{R} = -7.950/(-45.110) = 0.176$

1259 $\hat{a} = -7.950, SE_a = 2.332$

1260 $\hat{b} = -45.110, SE_b = 1.043$

1261 $t = 2.035$, the upper 2.5% point of a t -distribution with 33 (=40 - 7) degrees of freedom (for a 95%
 1262 confidence interval)

1263 $g = (2.035 * 1.043/(-45.110))^2 = 0.0022$.

1264
 1265 The 95% confidence interval can now be calculated:

1266

$$\frac{0.176 \pm \frac{2.035}{-45.110} \sqrt{(1 - .0022)(2.332)^2 + (0.176)^2 (1.043)^2}}{1 - .0022} = (0.071, 0.282)$$

1267

1268 Because the dilutions were coded as integers of the 1.5 dilutions, we now take base 1.5 antilogs. The
1269 (uncorrected for dilution) relative potency is $1.5^{0.176} = 1.074$ with a 95% confidence interval of

1270 (1.029, 1.121). A correction factor of $\frac{670 * 16.7 / 25}{20,000 * 1 / 40} = 0.89512$ is necessary because the dilutions

1271 were not exactly equipotent on the basis of the assumed potency. Multiplying by this correction
1272 factor and the assumed potency of 20,000 IU/vial yields a potency of 19,227 IU/vial with 95 per
1273 cent confidence limits from 18,422 to 20,069 IU/vial.

1274

1275 **7.3 Sigmoid Analysis for Continuous Data.** The data are from a macrophage cell lysis (MCL)
1276 assay used to determine potency of recombinant Protective Antigen 102 (rPA102). Standard and
1277 Test material were prepared at 12 (approximately) 2- to 3-fold dilutions, from 150 ng/mL to 0.8
1278 ng/mL rPA. Luminescence was measured at each dilution. The data are presented in Table 7.3.1 and
1279 Figure 7.3.1.

1280

1281

Table 7.3.1 Macrophage Cell Lysis Assay (MCL) Data

1282

rPA102 Conc. (ng/mL)		150	100	66.7	44.4	29.6	19.8
RLU	Test	164.1	187.4	208.1	268.8	687.7	3246.9
	Standard	172.1	171.4	224.8	337.5	955.8	2903.4
rPA102 Conc. (ng/mL)		12.2	8.8	3.9	2.6	1.7	0.8
RLU	Test	8442.7	12405.3	13764.7	13684.6	14339.6	13287.8
	Standard	9027	13657.3	14511	14784.5	14670.4	13738

1283

1284 These data will be used as an example of the four-parameter logistic model of Section 5.3.4.

1285

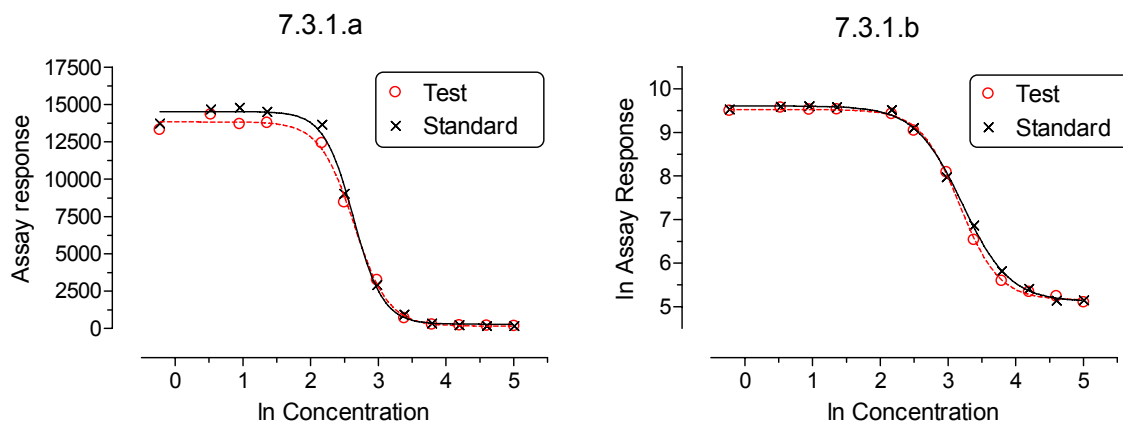
1286 *Step 1.* Examine the data and determine if there is need for a transformation of the response prior to
1287 analysis.

1288

1289 The left panel of Figure 7.3.1 shows a plot of the response as measured and shown in Table 7.1
1290 against the (natural) log of concentration. The right panel of Figure 7.1 shows the same data except
1291 that the response has been transformed to the natural log scale as well. The plotted curves are
1292 determined without assuming parallelism. The curves on the right show the characteristic symmetry
1293 about the center of the four-parameter curve somewhat better than do the curves on the left. Based
1294 on these plots, analysts may choose to analyze these data in the log-response scale. The difference is
1295 not large here, and some analysts may choose not to log transform the response. There is also an
1296 assumption made here regarding the difference in upper asymptotes seen in Figure 7.3.1a but not as
1297 evident in 7.3.1b, namely that the difference is within the system suitability range for the upper
1298 asymptote comparison.

1299

1300 **Figure 7.3.1. Plot of Response (Left) and ln Response (Right) against ln Concentration**



1301
1302

1303 *Step 2.* Assess suitability parameters on the Standard sample.

1304
1305 As part of making a decision about whether or not an assay yielded acceptable results, some
1306 laboratories test to determine if the Standard sample dose–response curve conforms acceptably to
1307 expectations based on previous assay experience. Statistics for the Standard curve are compared to
1308 recorded historical assay behavior. Such system suitability parameters are obtained in the curve
1309 fitting and include A , B , $\ln C$, D , and $RMSE$.

1310
1311 *Step 3.* Determine whether weighting is required and the weights, if needed.

1312
1313 Not shown here are data from other runs of this assay that indicate that the coefficient of variation
1314 is constant across concentration when measured in the original scale. Because the coefficient of
1315 variation is constant, the standard deviation is then proportional to the average response and
1316 increases as concentration increases. An analyst who chooses to analyze these data in the original
1317 scale must then properly weight the analyses in order for the resulting standard errors and
1318 confidence intervals to be correct.

1319
1320 A property of the log transform is that it converts constant coefficient of variation in the original
1321 scale into constant standard deviation in the log-transformed scale. Thus, if data are analyzed in the
1322 log scale, as will be done here, no weighting is required.

1323
1324 *Step 4.* Test for parallelism.

1325
1326 The model for a four-parameter logistic curve [5.6] is

1327
$$y = D + \frac{A - D}{1 + \exp[M(\ln z - b)]}$$

1328 where y is the natural log of the response and x is concentration.

1329
1330 The test for parallelism requires a model that combines Standard and Test data and accommodates
1331 Standard and Test data that have different asymptotes and slopes and provides an estimate of

1332 relative potency. Combining in this manner allows straightforward inference about the Delta terms
 1333 because the standard errors are then part of analysis output. The requisite function is:
 1334

1335
$$y = (D + T * \Delta D) + \frac{(A + T * \Delta A) - (D + T * \Delta D)}{1 + e^{(M + T * \Delta M) * (\ln z - b + T * \ln \rho)}}$$

1336 where T is an indicator variable that takes the value 0 for a Standard observation and 1 for a Test
 1337 observation, and ρ is the test material's relative potency.
 1338
 1339

1340 **Table 7.3.3. Estimates of Parallelism Parameters**

Parameter	Estimate	Standard		
		Error	90% CI	90% CI as % of $A-D$
ΔA	-0.0816	0.0447	(-0.0032, 0.16)	(-0.05, 2.5)
ΔD	0.0420	0.0647	(-0.0714, 0.155)	(-1.1, 2.4)
ΔM	0.5779	0.2168	(-0.198, 0.958)	(-6, 30)

1341
 1342 The 90% CI endpoints for the Deltas for A , B , and M are simply $\pm t_{0.95, 15}$ (95th percentile of a t -
 1343 distribution with 15 degrees of freedom) times the standard error of each Delta estimate. A 90% CI
 1344 is used rather than 95% because the objective is to show equivalence to 0. A 90% CI corresponds to
 1345 a 5% test of the equivalence hypothesis. The 90% CI endpoints shown as percentages (rightmost
 1346 column) are the 90% CI endpoints divided by the Standard range ($A-D$) for the asymptotes or by
 1347 the Standard slope for ΔM . These 90% CI endpoints shown as percentages can be compared to
 1348 sample acceptance criteria of $\pm 5-15\%$ for asymptotes and $\pm 35-50\%$ for slope. As shown in
 1349 Example 7.2 these acceptance criteria are examples and are not intended to apply to all assays. The
 1350 choice of criteria will depend on the assay and product. Generally, products with narrow therapeutic
 1351 windows generally should use the narrower end of these acceptance limits, but vaccines or other
 1352 products with wide therapeutic windows can use the wide range of these acceptance limits.
 1353

1354 *Step 5.* Refit the model assuming parallelism.

1355
 1356 The results provided in Table 7.3.3 support a conclusion of acceptable parallelism, although the
 1357 difference in slopes has a wide confidence interval. The next step is to fit the model assuming
 1358 common slope and intercepts (i.e., model [5.9]). It would not be appropriate to compare the EC50s
 1359 from the curves determined separately for the Test and Standard. The following model may now be
 1360 fit to determine parameter and relative potency estimates:
 1361

1362
$$y = D + \frac{A - D}{1 + e^{M * (\ln x - b - T * \ln \rho)}}$$

1364

Table 7.3.4. Final Parameter Estimates

Parameter	Estimate	Standard Error	95% CI
<i>A</i>	9.5700	0.0256	(9.5164, 9.6236)
<i>D</i>	5.1382	0.0369	(5.0610, 5.2154)
<i>M</i>	3.1135	0.1206	(2.8611, 3.3660)
<i>b</i>	3.2159	0.0176	(3.1790, 3.2529)
<i>lnρ</i>	0.0499	0.0216	(.0048, .0951)

1365

1366 The estimate of the relative potency, ρ (%), is $100 * e^{0.0499} = 105.1\%$; the 95% CI for ρ is $(100 * e^{0.0048}\%$,
 1367 $100 * e^{0.0951}\%)$, or (100.5%, 110.0%).

1368

1369 **7.4 Slope Ratio Model.** A laboratory carries out a chromogenic assay of factor VIII activity. Eight
 1370 replicates (more than might be used routinely) at 3 dilution points (IU/mL) are prepared for both a
 1371 Standard and a Test preparation. Additionally, 8 replicates of a blank are prepared. The data are
 1372 presented in Table 7.4.1 and are depicted graphically in Figure 7.4.1. As is seen in this Figure, there
 1373 is a loss in linearity with inclusion of the Blank data; therefore, the Blank responses will not be used
 1374 in calculations. However, the evident linearity across dilution points and similarity of intercepts
 1375 provide assurance that the slope-ratio model is appropriate for analysis of these data.

1376

1377

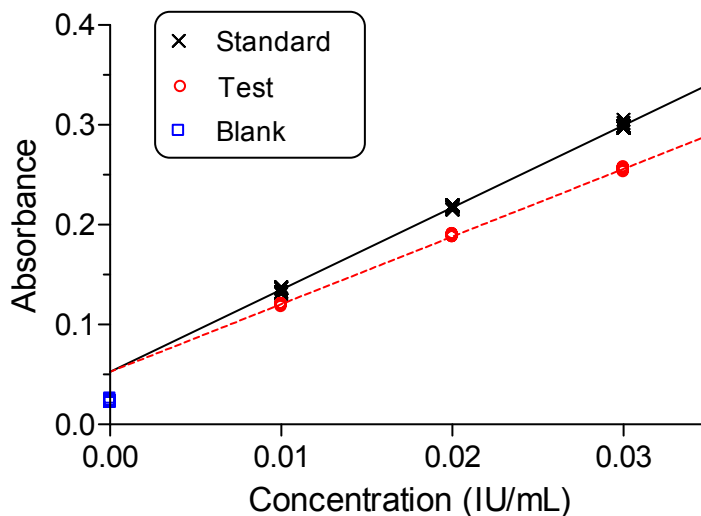
Table 7.4.1. Factor VIII Assay Data

		Concentration (IU/mL)					
		Standard			Test		
Absorbance	Blank	0.01	0.02	0.03	0.01	0.02	0.03
	0.022	0.133	0.215	0.299	0.120	0.188	0.254
	0.024	0.133	0.215	0.299	0.119	0.188	0.253
	0.024	0.131	0.216	0.299	0.118	0.190	0.255
	0.026	0.136	0.218	0.297	0.120	0.190	0.258
	0.023	0.137	0.220	0.297	0.120	0.190	0.257
	0.022	0.136	0.220	0.305	0.121	0.191	0.257
	0.022	0.138	0.219	0.299	0.121	0.191	0.255
	0.023	0.137	0.218	0.302	0.121	0.190	0.254

1378

1379

Figure 7.4.1. Slope Ratio Data



1380
1381
1382
1383
1384
1385
1386
1387
1388
1389

The first task is to determine whether the Test and Standard lines have a common intercept. Note that the intercept need not be 0. To test this, a linear regression that has separate slopes and intercepts for the Test and Standard and a difference of intercepts is fit. Table 7.4.2 shows the data of Table 7.4.1 formatted to facilitate this calculation. The column variable *Test* is set to 1 for Test data and 0 for Standard data; the coefficient for this variable will be the difference of intercepts. The other two columns enable calculation of slope coefficients for the Test and Standard regressions.

Table 7.4.2. Factor VIII Assay Data Formatted for Analysis

Absorbance	Conc_Std	Conc_Test	Test
0.133	0.01	0	0
0.133	0.01	0	0
0.131	0.01	0	0
0.136	0.01	0	0
0.137	0.01	0	0
0.136	0.01	0	0
0.138	0.01	0	0
0.137	0.01	0	0
0.215	0.02	0	0
0.215	0.02	0	0
0.216	0.02	0	0
0.218	0.02	0	0
0.220	0.02	0	0
0.220	0.02	0	0
0.219	0.02	0	0
0.218	0.02	0	0
0.299	0.03	0	0
0.299	0.03	0	0
0.299	0.03	0	0

Absorbance	Conc_Std	Conc_Test	Test
0.297	0.03	0	0
0.297	0.03	0	0
0.305	0.03	0	0
0.299	0.03	0	0
0.302	0.03	0	0
0.120	0	0.01	1
0.119	0	0.01	1
0.118	0	0.01	1
0.120	0	0.01	1
0.120	0	0.01	1
0.121	0	0.01	1
0.121	0	0.01	1
0.121	0	0.01	1
0.188	0	0.02	1
0.188	0	0.02	1
0.190	0	0.02	1
0.190	0	0.02	1
0.190	0	0.02	1
0.191	0	0.02	1
0.191	0	0.02	1
0.190	0	0.02	1
0.254	0	0.03	1
0.253	0	0.03	1
0.255	0	0.03	1
0.258	0	0.03	1
0.257	0	0.03	1
0.257	0	0.03	1
0.255	0	0.03	1
0.254	0	0.03	1

1390
1391 Results from the fit of this regression are given in Table 7.4.3. Regression lines are plotted in Figure
1392 7.4.1.
1393
1394

Table 7.4.3. Regression Results with Unequal Intercepts

	Coefficients	Standard Errors	<i>t</i> Statistics	<i>p</i> -values	Lower 95% CI	Upper 95% CI
Intercept	0.0530	0.0011	47.8	<<0.001	0.0507	0.0552
Difference of Intercepts	4.2×10^{-5}	0.0016	0.027	0.979	-0.0031	0.0032
Standard Slope	8.225	0.0513	160.4	<<0.001	8.122	8.328
Test Slope	6.769	0.0513	132.0	<<0.001	6.665	6.872

1395
1396 The estimated difference of intercepts is very small. Also, the confidence interval for the difference
1397 of the intercepts, (-0.0031, 0.0032), spans a range of values, all of which are negligible. This supports
1398 a choice to proceed with analysis as though there is a common intercept. If a more elaborate analysis

1399 of difference of intercepts is desired, one based on assessing equivalence with a prespecified
 1400 equivalence interval is recommended.

1401
 1402 Refitting the lines with a common intercept (i.e., not using the rightmost column of Table 7.4.2)
 1403 yields a new set of regression results, several of which are shown in Table 7.4.4. Computing the CI
 1404 for the relative potency using Fieller's Theorem requires the interslope covariance (*CO* in the
 1405 notation of Section 4.10).

1406
 1407 **Table 7.4.4. Regression Results with Equal Intercepts**

	Coefficients	Standard Errors
Intercept	0.0530	0.000774
Standard Slope	8.224	0.0383
Test Slope	6.770	0.0383

1408
 1409 The relative potency is estimated as $6.770/8.224 = 0.823$. Using the notation of [4.10]:

1410 $\hat{R} = 0.823$

1411 $\hat{a} = 6.770, SE_a = 0.0383$

1412 $\hat{b} = 8.224, SE_b = 0.0383$

1413 $C = 0.0011$

1414 $t = 2.014$, the upper 2.5% point of a *t*-distribution with 45 (= 48 - 3) degrees of freedom (for a 95%
 1415 CI).

1416
 1417 Then *g* is calculated to be $(2.014 * 0.0383/8.224)^2 = 8.8 \times 10^{-5}$, which will be taken as 0. Thus, $1 - g =$
 1418 1 in the Fieller's theorem formula. The CI is then calculated as

1419
$$CI = 0.823 \pm \frac{2.014}{8.224} \sqrt{(0.0383)^2 + (0.823)^2 (0.0383)^2 - 2 * 0.823 * 0.0011} = (0.817, 0.829)$$

1420 Finally a relative potency, as percent, of 82.3% with a 95% CI of (81.7%, 82.9%) is found.

1421
 1422 **7.5. Probit Analysis of Quantal Data.** The following data are from an animal assay. At each of the
 1423 analyte concentrations ($\mu\text{g/mL}$), 10 animals were tested for both Standard and Test preparations.
 1424 Each animal could respond (value = 1) or not respond (value = 0). The response for this assay is
 1425 thus dichotomous (quantal), and the data may be used as to demonstrate probit analysis. Although
 1426 this is not the logit analysis shown in Section 5.4, the two methods are similar in process.

1427
 1428 Response counts at each concentration are tabulated in Table 7.5.1. Figure 7.5.1 shows the
 1429 proportion responding plotted against the log of concentration. Notable features of these data are
 1430 the large number of cases with 90% and 100% responding and the steepness of the response curve
 1431 and the paucity of information at the center of the curve.

1432

1433

Table 7.5.1. Count of Positive Animal Responses

Concentration	Standard	Test	Concentration	Standard	Test
0.08	0	0	2.56	9	10
0.16	1	2	5.12	10	10
0.32	7	3	10.24	10	10
0.64	10	6	20.48	10	10
1.28	9	10			

1434

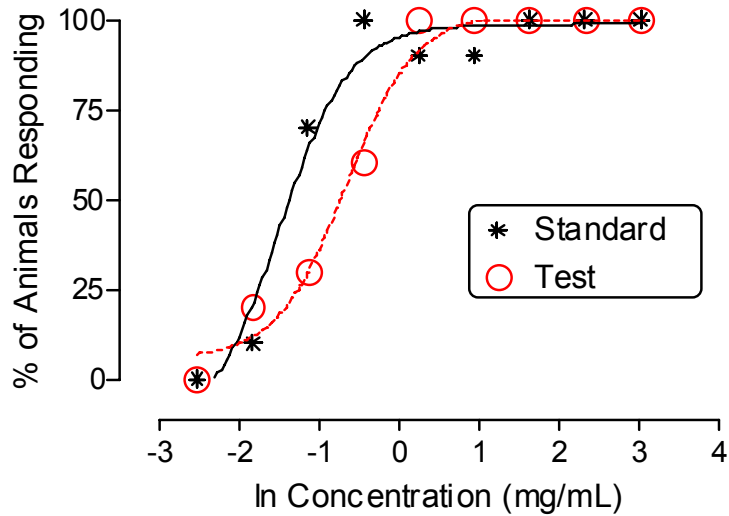
1435

1436

**Figure 7.5.1. Proportion Responding by Concentration
(Curves Shown Are Fitted Probit Curves)**

1437

1438



1439

1440

Following is the probit analysis:

1442

Step 1. Check for parallelism.

1444

We fit a probit model using three independent variables. For determining the difference of slopes, the three variables are T ($= 1$ for Test and 0 for Standard), $\log(\text{concentration})$, and $T \cdot \log(\text{concentration})$.

The parameter estimate for $\log(\text{concentration})$ is the slope for the Standard. The parameter estimate for

$T \cdot \log(\text{concentration})$ is the difference of slopes. Results are (using base 10 logarithms):

1449

1450

Table 7.5.2. Results for Difference of Slopes

Standard Slope	Difference of Slopes (SE)	90% CI for Difference of Slopes	Difference as a % of Standard Slope	90% CI as a % of Standard Slope
2.536	0.569 (0.837)	-0.807, 1.946	22.4%	-31.8%, 76.7%

1451

1452 The 90% CI for the difference of slopes is found as $0.569 \pm 1.645 \cdot 0.837$. This CI may be expressed
 1453 in terms of percentages of the Standard slope: $= 100\% \cdot (-0.807/2.536)$, $100\% \cdot (1.946/2.536) = -$
 1454 31.8% , 76.7% . Note that the CI so expressed does not represent a CI for (Difference of
 1455 slopes/Standard slope), a calculation that would require use of Fieller's Theorem.

1456
 1457 For determining the ratio of the Test slope to the Standard slope, an alternative choice of
 1458 independent variables simplifies calculations $-T$, $T \cdot \log(\text{concentration})$, and $(1 - T) \cdot \log(\text{concentration})$. With
 1459 these variables, the parameter estimate for $T \cdot \log(\text{concentration})$ is now the slope for the Test. The
 1460 parameter estimate for $(1 - T) \cdot \log(\text{concentration})$ is the slope for the Standard. With this choice of
 1461 variables we can easily obtain standard errors and use Fieller's Theorem to obtain a CI for the ratio.
 1462 With this choice results are

1463
 1464 Table 7.5.3. Results for Ratio of Slopes

Standard Slope (SE)	Test Slope (SE)	Ratio of Slopes	90% CI for Ratio of Slopes
2.536 (0.505)	3.105 (0.667)	1.224	0.730, 2.013

1465
 1466 In the Fieller's calculations, $CO = 0$. The two estimates are uncorrelated because this approach only
 1467 fits the two separate curves simultaneously. For probit and logit analyses, the t -value is replaced by a
 1468 standard normal value, 1.645 in this case.

1469
 1470 Although the difference and ratio of slopes is not large, the CIs are quite wide. We cannot be
 1471 confident that the two curves are parallel. In the data (Table 7.5.1), we see that the Test is somewhat
 1472 less potent (fewer responders) in the middle of the curve (concentrations of 0.32 and 0.64), but it is
 1473 more potent at the upper asymptote because it consistently obtains 100% response earlier than does
 1474 the Standard.

1475
 1476 For purposes of illustrating the relative potency calculation we will proceed as if the two curves are
 1477 accepted as sufficiently parallel.

1478
 1479 *Step 2.* Determine the estimate of relative potency and its CI.

1480
 1481 To determine an estimate of the relative potency, we refit the probit model, now with two
 1482 independent variables, T and $\log(\text{concentration})$. The difference of intercepts is the parameter
 1483 corresponding to T .

1484
 1485 Table 7.5.4. Results for Relative Potency

Difference of Intercepts (SE)	Common Slope (SE)	Log ₁₀ of Relative Potency (90% CI)	% Relative Potency (90% CI)
-0.313 (0.383)	2.771 (0.492)	-0.113 (-0.359, 0.118)	77.1% (43.7%, 131.4%)

1486
 1487 For Fieller's Theorem, we need
 1488 $a = -0.313$ $SE_a = 0.383$
 1489 $b = 2.771$ $SE_b = 0.492$
 1490 $CO = -0.009$ (from output; not shown in table)
 1491 $t = 1.645$

1492
1493
1494
1495

The g factor is then $(1.645 \cdot 0.492 / 2.771)^2 = 0.085$. The square root term in the formula for CI for the log of the relative potency (a/b) is then

1496
$$(-0.359, 0.118) = \frac{\left\{ -0.113 - \frac{0.085 \cdot (-0.009)}{(0.492)^2} \pm \frac{1.645}{2.771} \sqrt{(1 - 0.085)(0.383)^2 + (-0.113)^2(0.492)^2 - 2 \cdot (-0.113) \cdot (-0.009) + \frac{0.085 \cdot (-0.009)^2}{(0.492)^2}} \right\}}{1 - 0.085}$$

1497
1498
1499
1500

The estimate of relative potency and its CI is then found by taking base10 antilogs of the log values, e.g., $77.1\% = 100\% \cdot 10^{-0.113}$. We see that the CI for the relative potency is quite wide.

1501

8.0 LITERATURE

1502

1503 Bliss CI. *The Statistics of Bioassay*. Academic Press, New York; 1952.

1504

1505 Bliss CI. Analysis of the biological assays in U.S.P. XV. *Drug Standards*. 1956;24:33–67.

1506

1507

1508 Brown F, Mire-Sluis A, eds. *The Design and Analysis of Potency Assays for Biotechnology Products*. Karger, New York; 2002.

1509

1510

1511 DeLean A, Munson PJ, Rodbard D. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves, *Am J Physiol*. 1978;235:E97–E102.

1512

1513

1514 European Directorate for the Quality of Medicines. *European Pharmacopoeia*, Chapter 5.3, Statistical Analysis. EDQM, Strasburg, France; 2004:473–507.

1515

1516

1517 Finney DJ. *Probit Analysis*, 3rd ed. Cambridge, Cambridge University Press; 1971.

1518

1519

1520 Finney DJ. *Statistical Method in Biological Assay*. 3rd ed. London, Griffin; 1978.

1521

1522 Govindarajulu Z. *Statistical Techniques in Bioassay*. 2nd ed. New York, Karger; 2001.

1523

1524

1525 Hauck WW, Capen RC, Callahan JD, et al. Assessing parallelism prior to determining relative potency. *PDA J Pharm Sci Technol*. 2005;59:127–137.

1526

1527

1528 Hewitt W. *Microbiological Assay for Pharmaceutical Analysis: A Rational Approach*. New York, Interpharm/CRC, 2004.

1529

1530

1531 Higgins KM, Davidian M, Chew G, Burge H. The effect of serial dilution error on calibration inference in immunoassay. *Biometrics* 1998;54:19–32.

1532

1533

1534 Hurlber, SH. Pseudo replication and the design of ecological field experiments. *Ecological Monogr*. 1984;54:187–211.

1535

1536

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

- 1537 Nelder JA, Wedderburn RWM. Generalized linear models, *J Royal Statistical Soc, Series A*.
1538 1972;135:370–384.
1539
1540 Rorabacher DB. Statistical treatment for rejection of deviant values: critical values of Dixon's "Q"
1541 parameter and related subrange ratios at the 95% confidence level. *Anal Chem* 1991;63:39–48.