

<1030> BIOLOGICAL ASSAY CHAPTERS—OVERVIEW AND GLOSSARY

USP–NF contains four general chapters regarding the development, validation, and analysis of bioassays (biological assays): *Design and Analysis of Biological Assays* <111>, *Design and Development of Biological Assays* <1032>, *Biological Assay Validation* <1033>, and *Analysis of Biological Assays* <1034>. This proposed new chapter, *Biological Assay Chapters—Overview and Glossary* <1030>, provides an overview and some material common to chapters <1032>, <1033>, and <1034>, including a glossary of bioassay-related terms.

The suite of *USP* bioassay chapters focuses on relative potency assays. These assays recognize the inherent variability in biological test systems (whether animals or cells) that may be seen from laboratory to laboratory and from day to day. That inherent variability compromises the reliability of an absolute measure of potency. In relative potency assays, the biological activity of a Test material is compared to the activity of a Standard in an assay system wherein the use of a Standard reduces the influence of the inherent variability of the system on the estimation of relative potency. Relative potency assays also provide focus on important variability in response because of differences between the Test and Standard materials (if such a difference exists). The Test is expected to behave as a dilution or concentration of the Standard and should exhibit the property of *similarity*. Although they are intended for relative potency bioassays, the principles and practices developed in these chapters may have wider application—for example, to immunoassays and receptor-ligand–binding assays used to determine relative potency.

Chapter <1032> provides information for scientists developing a new biological assay. As seen in *Table 1*, the chapter covers a range of activities across the life cycle of the assay, with emphasis on development leading to validation, including the choice of test system and design considerations (e.g., plate layout). It also addresses data analysis strategies that should be considered during development (before validation) but that are not routinely addressed later. Among these strategies are the choice of weighting scheme, data transformation, if any, and choice of statistical model. Statistical details in support of these sections of chapter <1032> are found in chapter <1034>.

Table 1. Primary Sections of *Design and Development of Biological Assays* <1032>

Section	Section Title
1	Introduction
1.1	Purpose and Scope
1.2	Audience
2	Bioassay Fitness for Use
2.1	Process Development
2.2	Process Characterization
2.3	Product Release
2.4	Process Intermediates
2.5	Stability
2.6	Qualification of Reagents
2.7	Product Integrity
3	Bioassay Fundamentals
3.1	In Vivo Bioassays
3.2	Ex Vivo Bioassays
3.3	In Vitro (Cell-Based) Bioassays
3.4	Standard
4	Statistical Aspects of Bioassay Fundamentals
4.1	Data
4.2	Assumptions
4.3	Variance Heterogeneity, Weighting, and Transformation
4.4	Normality
4.5	Linearity of Concentration–Response Data
4.6	Common Bioassay Models
4.7	Suitability Testing
4.8	Outliers
4.9	Fixed and Random Effects in Models of Bioassay Response
5	Stages in the Bioassay Development Process

Table 1. Primary Sections of *Design and Development of Biological Assays* (1032) (continued)

Section	Section Title
5.1	Design: Assay Layout, Blocking, and Randomization
5.2	Development
5.3	Data Analysis during Assay Development
5.4	Bioassay Validation
5.5	Bioassay Maintenance

Chapter (1034) provides information about the data analyses appropriate for common relative potency bioassays, including parallel-line, slope-ratio, parallel-curve, and quantal models. The chapter also includes analyses supporting system and sample suitability assessment and methods for combining results from independent assays (see *Table 2*). This chapter is the most statistically advanced of the three chapters but is designed to be suitable for both biologists and statisticians. The conceptual material requires only a minimal statistical background. Methods sections require a statistical background at the level of *Analytical Data—Interpretation and Treatment* (1010) and familiarity with linear regression.

Table 2. Primary Sections of *Analysis of Biological Assays* (1034)

Section	Section Title
1	Introduction
2	Overview of Analysis of Bioassay Data
3	Analysis Models
3.1	Quantitative and Qualitative Assay Responses
3.2	Overview of Models for Quantitative Responses
3.3	Parallel-Line Models for Quantitative Responses
3.4	Nonlinear Models for Quantitative Responses
3.5	Slope–Ratio Concentration–Response Models
3.6	Dichotomous (Quantal) Assays
4	Confidence Intervals
4.1	Combining Results from Multiple Assays
4.2	Combining Independent Assays (Sample-Based Confidence Interval Methods)
4.3	Model-Based Methods
5	Additional Sources of Information

Chapter (1033) is intended to follow chapters (1032) (assay development) and (1034) (development of data analysis plans). That is, chapter (1033) assumes a fully developed bioassay (including a data analysis plan and at least tentative values for system and sample suitability criteria and the bioassay format) and provides guidance about the validation of that assay. The chapter addresses the validation characteristics relevant to relative potency bioassays and provides more detail regarding the statistical methods used in validation than does *Validation of Compendial Procedures* (1225). Principles and practices developed in chapter (1033), although they are intended for relative potency assays, may have wider application. The chapter emphasizes validation approaches that provide flexibility in adopting new bioassay methods, new biological drug products, or both (see *Table 3*).

Table 3. Primary Sections of *Biological Assay Validation* (1033)

Section	Section Title
1	Introduction
2	Fundamentals of Bioassay Validation
2.1	Bioassay Validation Protocol
2.2	Documentation of Bioassay Validation Results
2.3	Bioassay Validation Design
2.4	Validation Strategies for Bioassay Performance Characteristics
2.5	Validation Target Acceptance Criteria
2.6	Assay Maintenance
2.7	Statistical Considerations
3	A Bioassay Validation Example

Table 3. Primary Sections of *Biological Assay Validation* (1033) (continued)

Section	Section Title
3.1	Intermediate Precision
3.2	Relative Accuracy
3.3	Range
3.4	Use of Validation Results for Bioassay Characterization
3.5	Confirmation of Intermediate Precision and Revalidation
4	Additional Sources of Information
Appendix	Measures of Location and Spread for Lognormally Distributed Variables

GLOSSARY

This glossary pertains to biological assays and provides a compendial perspective that is consistent across *USP–NF*'s suite of bioassay chapters, is complementary to previous authoritative usage, and provides a useful focus on the bioassay context. In many cases the terms cited here have common, though undocumented, usages or are defined in *Validation of Compendial Procedures* (1225) and in the International Conference on Harmonization (ICH) Guideline Q2(R1), *Validation of Analytical Procedures: Text and Methodology*.¹ (Chapter (1225) and ICH Q2(R1) agree on definitions.) The *Glossary* is intended to be consistent with these precedent usages, and notes are provided when a difference arises because of the bioassay context. Definitions from (1225) and ICH Q2(R1) are identified as, for example, “(1225)” if taken without modification, or “adapted from (1225)” if taken with minor modification for application to bioassay. Most definitions are accompanied by notes that elaborate on the bioassay context.

The terms are organized alphabetically within five topic sections:

- I. General terms related to bioassays
- II. Terms related to performing a bioassay
- III. Terms related to precision and accuracy
- IV. Terms related to validation
- V. Terms related to statistical design and analysis

Table 4 shows each term and the *Glossary* section in which it can be found.

Table 4. Terms Listed in the *Glossary*

Term	Section	Term	Section
Accuracy	III	Mixed-effects model	V
Analysis of variance (ANOVA)	V	Modeling, statistical	V
Analytical procedure	I	Nested	V
Assay	I	Out of specification	II
Assay data set	I	Parallelism	V
Bioassay	I	Partially crossed	V
Biological assay	I	Point estimate	V
Blocking	V	Potency	I
Complete block design	V	Precision	III
Confidence interval	V	Pseudoreplication	V
Crossed	V	<i>P</i> value	V
Design of experiments (DOE)	V	Quantitation limit	IV
Dilutional linearity	IV	Random effect	V
Direct bioassays	I	Random error	III
Equivalence test	V	Random factor	V
Errors, types of	III	Randomization	V
Expected mean square	V	Range	IV
Experimental design	V	Relative bias	III
Experimental unit	V	Relative potency	I

¹ Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1_Guideline.pdf. Accessed 29 March 2012.

Table 4. Terms Listed in the *Glossary* (continued)

Term	Section	Term	Section
Factor	V	Repeatability	III
Factorial design	V	Replication	V
Fixed effect	V	Reportable value	I
Fixed factor	V	Reproducibility	III
Format variability	III	Robustness	IV
Format, bioassay	II	Run	I
Fractional factorial design	V	Sample suitability	II
Full factorial design	V	Significance probability	V
General linear model	V	Similar preparations	I
Geometric coefficient of variation	III	Similarity (algebraic)	I
Geometric standard deviation	III	Specificity	III
Incomplete block design	V	Standard error of estimate	V
Independence	V	Statistical process control (SPC)	V
Indirect bioassays	I	System suitability	II
Interaction	V	Systematic error	III
Intermediate precision	III	True replicates	V
Level	V	Truncation bias	III
Linearity, dilutional	IV	Type I error	V
Lognormal distribution	V	Type II error	V
Lower limit of quantitation	IV	Validation, assay	IV
Mean square	V	Variance component analysis	V

I. General Terms Related to Bioassays

ANALYTICAL PROCEDURE [ADAPTED FROM Q2(R1)]

Detailed description of the steps necessary to perform the analysis.

[NOTE—1. The procedure may include but is not limited to the sample preparation, the Reference Standard, and the reagents; use of equipment; generation of the standard curve; use of the formulae for the calculation; etc. 2. An FDA Guidance² provides a list of information that typically should be included in the description of an analytical procedure.]

ASSAY

Analytical procedure to determine the quantity of one or more components or the presence or absence of one or more components.

[NOTE—1. *Assay* often is used as a verb synonymous with *test* or *evaluate*, as in “I will assay the material for impurities.” In this glossary, *assay* is a noun and is synonymous with *analytical procedure* (q.v.). 2. The phrase *to run the assay* means to perform the analytical procedure as specified. 3. In common practice, *assay* and *run* (q.v.) often are used interchangeably. In this glossary, they are different. Also see *bioassay* and *bioassay data set*.]

ASSAY DATA SET

The set of data used to determine a single potency or relative potency for all samples included in the bioassay.

[NOTE—1. The definition of an assay data set can be subject to interpretation as necessarily a minimal set. It may be possible to determine a potency or relative potency from a set of data but not do this well. It is not the intent of this definition to mean that an assay data set is the minimal set of data that can be used to determine a relative potency. In practice, an assay data set should include, at least, sufficient data to assess similarity (q.v.). It also may include sufficient data to assess other assumptions. 2. It is also not an implication of this definition that assay data sets used together in determining a reportable value (q.v.) are necessarily independent from one another, although it may be desirable that they be so. When a run (q.v.) consists of multiple assay data sets, independence of assay sets within the run must be evaluated.]

² Analytical Procedures and Methods Validation for Drugs and Biologics, Guidance for Industry. 2015. <http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm386366.pdf>. Accessed 20 April 2016.

BIOASSAY, BIOLOGICAL ASSAY (these terms are interchangeable)

Analysis (as of a drug) to quantify the biological activity or activities of one or more components by determining its capacity for producing an expected biological activity on a culture of living cells (in vitro) or on test organisms (in vivo), expressed in terms of units.

[NOTE—1. The components of a bioassay include the analytical procedure, the statistical design for collecting data, and the method of statistical analysis that eventually yields the estimated potency or relative potency. 2. Bioassays can be either direct or indirect.

Direct bioassays—Bioassays that measure the concentration of a substance that is required in order to elicit a specific response. For example, the potency of digitalis can be directly estimated from the concentration required to stop a cat's heart. In a direct assay, the response must be distinct and unambiguous. The substance must be administered in such a manner that the exact amount (threshold concentration) needed to elicit a response can be readily measured and recorded.

Indirect bioassays—Bioassays that compare the magnitude of responses for nominally equal concentrations of reference and test preparations rather than test and reference concentrations that are required to achieve a specified response. Most biological assays in *USP–NF* are indirect assays that are based on either quantitative or quantal (yes/no) responses.]

POTENCY [21 CFR 600.3(s)]

The specific ability or capacity of the product, as indicated by appropriate laboratory tests or by adequately controlled clinical data obtained through the administration of the product in the manner intended, to effect a given result.

[NOTE—1. A wholly impotent sample has no capacity to produce the expected specific response, as a potent sample would. Equipotent samples produce equal responses at equal dosages. Potency typically is measured relative to a Reference Standard or preparation that has been assigned a single unique value (e.g., 100.0) for the assay; see *relative potency*. At times, additional qualifiers are used to indicate the physical standard employed (e.g., "international units"). 2. Some biological products have multiple uses and multiple assays. For such products there may be different reference lots that do not have consistently ordered responses across a collection of different relevant assays. 3. [21 CFR 610.10] Tests for potency shall consist of either in vitro or in vivo tests, or both, which have been specifically designed for each product so as to indicate its potency in a manner adequate to satisfy the interpretation of potency given by the definition in 21 CFR 600.3(s).]

RELATIVE POTENCY

A measure obtained from the comparison of a Test to a Standard on the basis of capacity to produce the expected potency.

[NOTE—1. A frequently invoked perspective is that relative potency is the degree to which the Test preparation is diluted or concentrated relative to the Standard. 2. Relative potency is unitless and is given definition, for any test material, solely in relation to the reference material and the assay.]

REPORTABLE VALUE

The value that will be compared to an acceptance criterion.

[NOTE—1. The acceptance criterion for comparison may be in the *USP* monograph, or it may be set by the company, e.g., for product release. 2. The term *reportable value* is inextricably linked to the "intended use" of an analytical procedure. Assays are performed on samples in order to yield results that can be used to evaluate some parameter. Assays may have different summary values or formats for different purposes (e.g., lot release vs. calibration of a new reference standard). The reportable value may be different even if the mechanics of the test itself are identical. Validation is required in order to support the properties of each choice of reportable value. In practice there may be one physical document that is the analytical procedure used for more than one application, but each application must be detailed separately within that document. Alternatively, there may be separate documents for each application. 3. When the inherent variability of a biological response, or that of the log potency, precludes a single assay data set's attaining a value sufficiently accurate and precise to meet a specification, the assay format may be changed as necessary. The number of blocks or complete replicates needed depends on the assay's inherent accuracy and precision and on the intended use of the reported value. It is practical to improve the precision of a reported value by reporting the geometric mean potency from multiple assays. The number of assays used is determined by the relationship between the precision required for the intended use and the inherent precision of the assay system.]

RUN

The performance of the analytical procedure that can be expected to have consistent precision and trueness; usually, the assay work that can be accomplished by a single analyst in a set time with a given unique set of assay factors (e.g., standard preparations).

[NOTE—1. There is no necessary relationship of run to assay data set (q.v.). The term *run* is laboratory specific and relates to the laboratory's physical capability and environment for performing the work of an assay. An example of a run is given by one analyst's simultaneous assay of several samples in one day's bench work. During the course of a single run, it may be possible to determine multiple reportable values. Conversely, a single assay data set may include data from multiple runs. 2. From a statistical viewpoint, a run is one realization of the factors associated with intermediate precision (q.v.). Within-run variability is thus repeatability. It is good practice to associate runs with factors that are significant sources of variation in the assay. For example, if cell passage number is an important source of variation in the assay response obtained, then each change in cell passage number initiates a new run. If the variance associated with all factors that could be assigned to runs is negligible, then the influence of runs can be ignored in the analysis, and the analysis can focus on combining independent analysis data sets. 3. When a run contains multiple assays, caution is required regarding the independence of the assay results. Factors that typically

are associated with runs and that cause lack of independence include cell preparations, groups of animals, analyst, day, a common preparation of reference material, and analysis with other data from the same run. Even though a strict sense of independence may be violated because some elements are shared among the assay sets within a run, the degree to which independence is compromised may have negligible influence on the reportable values obtained and should be verified and monitored.]

SIMILAR PREPARATIONS

The property that the Test and the Standard contain the same effective constituent, or the same effective constituents in fixed proportions, and all other constituents are without effect in some specific assay context.

[NOTE—1. Having similar preparations is often summarized as the property that the Test behaves as a dilution (or concentration) of the Standard. 2. Similar preparations are fundamental to methods for determination of relative potency. Given similar preparations, a relative potency can be calculated, reported, and interpreted. In the absence of similar preparations, a meaningful relative potency cannot be reported or interpreted. 3. The practical consequence of similar preparations is algebraic similarity (q.v.). (Also see *Parallelism*, section V.)]

SIMILARITY (ALGEBRAIC)

The Test and Standard concentration-response curves are algebraically related in a manner consistent with similar preparations.

[NOTE—1. Examples of similarity are parallelism (q.v.) of concentration-response curves and equality of intercepts in slope ratio models. 2. Failure to statistically demonstrate dissimilarity between a Reference and a Test does not amount to demonstration of similarity. To demonstrate similarity an equivalence approach is appropriate; see <1032> and <1034>. 3. Similarity is typically a sample suitability (q.v.) criterion. Note, however, that suitability is a necessary but not sufficient condition for preparations to be similar. In practice, absent knowledge of differences between the Test and Standard materials, demonstration of similarity is accepted as demonstrating similar preparations.]

II. Terms Related to Performing a Bioassay

FORMAT, BIOASSAY

The intra- and inter-run replication strategy for replication of assay data sets that has been determined by variance analysis to support the use of the bioassay.

[NOTE—1. Modifications to bioassay format may occur as new information regarding sources of variability becomes available. Such modifications do not include changes to the dilution scheme of Test samples or Standard, or the replication strategy (part of what is sometimes called bioassay configuration). Assay configuration can include nested dimensions like plate design, multiple plates per day, single plates on multiple days, etc. 2. The geometric mean relative potency determined from the bioassay format is the reportable value, which may be used to assess conformance to specifications or as a component of subsequent analysis (e.g., stability evaluation).]

OUT OF SPECIFICATION (OOS)

The property of a reportable value that falls outside its specification acceptance criterion.

[NOTE—*Out of specification* is not a property of the bioassay but rather a property of Test samples. The term is introduced into <1033> in conjunction with setting validation acceptance criteria which limit the risk of producing out-of-specification test results because of bioassay performance characteristics.]

SAMPLE SUITABILITY

A sample is suitable (may be used in the estimation of potency) if its response curve satisfies limits on critical properties that are stated in the assay procedure.

[NOTE—Response curve properties are to be taken generally; i.e., includes outliers and variability. The most significant of these properties for bioassays is similarity (q.v.) to the standard response curve. In addition, all assay systems have limits on the range of values they can report. For samples that fail one or more sample suitability criteria in a bioassay, the potency estimate from those samples should not be used as a reportable value or as a contributor to a reportable value. Also see *truncation bias* in this *Glossary* and the sections *Sample Suitability* and *Range* in general chapter <1032>.]

SYSTEM SUITABILITY

An assay system is suitable for its intended purpose if it is capable of providing legitimate measurements as defined in the assay protocol.

[NOTE—System suitability may be thought of as an assessment of whether there is any evidence of a problem in the assay system. An example is provided by positive and negative controls, where values outside their normal ranges suggest that the assay system is not working properly.]

III. Terms Related to Precision and Accuracy

ACCURACY (1225)

The closeness of test results obtained by the procedure and the true value.

[NOTE—1. ICH and USP give the same definition of accuracy. However, ISO specifically regards accuracy as having two components, bias and precision.³ That is, to be *accurate* as used by ISO, a measurement must both be on target (have low bias) and be precise. In contrast, ICH Q2(R1) states that accuracy is sometimes termed “trueness” but does not define *trueness*. ISO defines *trueness* as the “closeness of agreement between the average value obtained from a large series of test results and an accepted reference value” and indicates that “trueness is usually expressed in terms of bias.” The 2001 FDA Guidance on Bioanalytical Method Validation⁴ defines *accuracy* in terms of “closeness of *mean* test results obtained by the method to the true value (concentration) of the analyte” (emphasis added) and thus is consistent with the ICH usage. This glossary adopts the USP/ICH approach. That is, *accuracy* is defined as the agreement between the mean (or expected results) from an assay and the true value, and uses the phrase *accurate and precise* to indicate low bias (accurate) and low variability (precise). 2. Considerable caution is needed when using or reading the term *accuracy*. In addition to the inconsistency between USP/ICH and ISO, common usage is not consistent. 3. For purposes of bioassay validation, the terms *accuracy* and *bias* have been replaced by *relative accuracy* and *relative bias*.]

ERROR, TYPES OF

Two sources of errors that affect the uncertainty of results of a biological assay are systematic error and random error.

A **systematic error** is one that happens with similar magnitude and consistent direction repeatedly. This introduces a bias in the determination. Effective experimental design, including randomization and/or blocking, can reduce systematic error.

A **random error** is one whose magnitude and direction vary without pattern. Random error is an inherent variability or uncertainty of the determination. Conversion of systematic into random error, through experimental design or randomization, increases the robustness of a biological assay and allows a comparatively simple analysis of assay data but may require a larger sample size.

FORMAT VARIABILITY

Predicted variability for a particular bioassay format.

GEOMETRIC COEFFICIENT OF VARIATION

Found as $\text{antilog}(S)-1$, where S is the standard deviation determined in the log scale.

[NOTE—The geometric coefficient of variation is usually reported as a percentage (%GCV). It is important not to confuse the %GCV with the %CV. The %GCV is a measure of spread relevant to data analyzed in the log-transformed [$Y = \log(X)$] scale, and the %CV is a measure relevant to data analyzed in the original (X) scale.]

GEOMETRIC STANDARD DEVIATION (GSD)

The variability of the log-transformed values of a lognormal response expressed as a percentage in the untransformed scale. It is found as $\text{antilog}(S)$, where S is the standard deviation determined in the log scale.

[NOTE—For example, if the standard deviation of log potency is S using log base 2, the GSD of potency is 100×2^S .]

INTERMEDIATE PRECISION (ADAPTED FROM (1225))

Within-laboratory precision associated with changes in operating conditions.

[NOTE—1. Factors contributing to intermediate precision involve anything that can change within a given laboratory and that may affect the assay, including different days, different analysts, different equipment, etc. Intermediate precision is thus “intermediate” in scope between the extremes of repeatability (intra-assay) and reproducibility (inter-laboratory). 2. Any statement of intermediate precision should identify the factors that varied. For example, “The intermediate precision associated with changing equipment and operators is....” 3. Investigators can benefit from separately identifying the precision associated with each source (e.g., inter-analyst precision). This may be part of assay development and validation when there is value in identifying the important contributors to intermediate precision. 4. When reporting intermediate precision, particularly for individual sources, care should be taken to distinguish between intermediate precision variance and components of that variance. The intermediate precision *variance* includes repeatability and thus must be at least as large as the repeatability variance. A *variance component*, e.g., associated with analyst, also is a part of the intermediate precision variance for analyst, but it could be negligible and need not be larger in magnitude than the repeatability variance.]

³ ISO. International Standard 5725-1. Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions. Geneva, Switzerland; 1994.

⁴ FDA. Guidance for Industry. Bioanalytical Method Validation. May 2001. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070107.pdf>. Accessed 7 December 2011.

PRECISION (〈1225〉)

Measure of agreement among individual test results when the procedure is applied repeatedly to multiple samplings of a homogeneous sample.

[NOTE—1. Precision may be considered at three levels: repeatability (q.v.), intermediate precision (q.v.), and reproducibility (q.v.). 2. Precision should be investigated using homogeneous, authentic samples. However, if it is not possible to obtain a homogeneous sample, precision can be investigated using spiked samples that mimic a true sample or a sample solution. 3. Precision may be expressed as the variance, standard deviation, coefficient of variation, or geometric coefficient of variation (q.v.).]

RELATIVE BIAS

Measure of difference between the expected (or mean) value and the true value, expressed as a percentage of the true value.

REPEATABILITY (〈1225〉)

The precision within a laboratory over a short interval of time, using the same analyst with the same equipment.

[NOTE—1. ICH Q2(R1) says that repeatability is also termed “intra-assay” precision. In the bioassay context, the better term is *intra-run*, and a “short interval of time” connotes *within-run*. 2. The idea of a “short interval of time” can be problematic with bioassays. If a run requires multiple weeks and consists of a single assay set, then intra-run precision cannot be determined. Alternatively, if a run consists of two assay data sets and a run can be done in a single day, repeatability of the relative potency determination can be assessed.]

REPRODUCIBILITY 〈1225〉

The precision between laboratories.

[NOTE—1. Reproducibility includes contributions from repeatability and all factors that contribute to intermediate precision, as well as any additional contributions from inter-laboratory differences. 2. Reproducibility applies to collaborative studies such as those for standardization or portability of methodology. Depending on the design of the collaborative study, it may be possible to separately describe variance components associated with intra- and inter-laboratory sources of variability.]

SPECIFICITY 〈1225〉

The ability to assess unequivocally the analyte in the presence of components that may be expected to be present.

[NOTE—1. Typically these components may include impurities, degradants, matrix, etc. See chapter 〈1225〉 for further discussion. 2. This definition is also associated with selectivity in other guidances for analytical methods. 3. Specificity can mean the measurement of the specific analyte of interest and no other similar analyte.]

TRUNCATION BIAS

Bias that occurs when some portion of the distribution of responses is not observed or recorded.

[NOTE—1. When there is truncation bias, the distribution of recorded observations does not match the true distribution of responses. 2. Truncation bias may occur in a bioassay that does not report estimates of log potency outside a set potency range. For example, a sample with a true potency at an edge of this range is expected to fail to yield (report) a potency estimate in approximately half of the assays in which it appears. In this example, the mean of the observed potencies will be biased toward log potency 0.]

IV. Terms Related to Validation

DILUTIONAL LINEARITY (ADAPTED FROM 〈1225〉)

The ability (within a given range) of a bioassay to obtain measured relative potencies that are directly proportional to the true relative potency of the samples.

[NOTE— 1. To determine dilutional linearity, sometimes called bioassay analytical linearity, across a range of known relative potency values, analysts examine the relationship between known log potency and mean observed log potency. If that relationship yields an essentially straight line with a y-intercept of 0 and a slope of 1, the assay has direct proportionality. If that plot yields an essentially straight line but either the y-intercept is not 0 or the slope is not 1, the assay has a proportional linear response. 2. To assess whether the slope is (near) 1.0 requires an a priori equivalence or indifference interval. It is not proper statistical practice to test the null hypothesis that the slope is 1.0 against the alternative that it is not 1.0 and then to conclude a slope of 1.0 if this is not rejected. Bioassay analytical linearity is separate from consideration of the shape of the concentration–response curve. Linearity of concentration–response is not a requirement of bioassay analytical linearity since bioassay analytical linearity is possible regardless of the form of the concentration–response curve. 3. Dilutional linearity is further addressed in 〈1033〉.]

QUANTITATION LIMIT (LOWER LIMIT OF QUANTITATION; ADAPTED FROM 〈1225〉)

The lowest known relative potency for which the assay has suitable precision and accuracy.

[NOTE—1. This applies to assay results (log potency) rather than the reportable value. 2. The quantitation limit is not commonly determined for relative potency bioassays. Animal assays with serologic endpoints are examples of the use of this term.]

RANGE (ADAPTED FROM <1225>)

The interval between the upper and lower known relative potencies (and including those relative potencies) for which the bioassay is demonstrated to have a suitable level of precision, accuracy, and bioassay analytical linearity.

[NOTE—This applies to reportable values (typically a geometric mean) rather than the individual assay results.]

ROBUSTNESS (<1225>)

A measure of an analytical procedure's capacity to remain unaffected by small but deliberate variations in method parameters listed in the procedure documentation.

[NOTE—1. Robustness is an indication of a bioassay's reliability during normal usage. For example, a cell culture assay system that is robust to the passage number of the cells can provide potency values with acceptable accuracy and precision across a consistent range of passage numbers. 2. ICH Q2(R1) states:

The evaluation of robustness should be considered during the development phase and depends on the type of procedure under study. It should show the reliability of an analysis with respect to deliberate variations in method parameters. If measurements are susceptible to variations in analytical conditions, the conditions should be suitably controlled or a precautionary statement should be included in the procedure. One consequence of the evaluation of robustness should be that a series of system suitability [q.v.] parameters (e.g., resolution test) is established to ensure that the validity of the analytical procedure is maintained whenever used.¹]

VALIDATION, ASSAY

Assay validation is the process of demonstrating and documenting that the performance characteristics of the procedure and its underlying method meet the requirements for the intended application and that the assay is thereby suitable for its intended use.

[NOTE—Formal validations are conducted prospectively according to a written plan that includes justifiable acceptance criteria on validation parameters. See <1033>.]

V. Terms Related to Statistical Design and Analysis

ANALYSIS OF VARIANCE (ANOVA)

A statistical tool used to assess contributions to variability from experimental factors.

BLOCKING

The grouping of related experimental units in experimental designs.

[NOTE—1. Blocking often is used to reduce the contribution to variability associated with a factor not of primary interest. 2. Blocks may consist, for example, of groups of animals (a cage, a litter, or a shipment), individual 96-well plates, sections of 96-well plates, or whole 96-well plates grouped by analyst, day, or batches of cells. 3. The goal is to isolate, by statistical design and analysis, a systemic effect, such as cage, so that it does not obscure the comparisons of interest.

A **complete block design** occurs when all levels of a treatment factor (in a bioassay, the primary treatment factors are sample and concentration) can be applied to experimental units for that factor within a single block. Note that the two treatment factors *sample* and *concentration* may have different experimental units. For example, if the animals within a cage are all assigned the same concentration but are assigned unique samples, then the experimental unit for concentration is cage and the experimental unit for sample is animal, and cage is a blocking factor for sample.

An **incomplete block design** occurs when the number of levels of a treatment factor exceeds the number of experimental units for that factor within the block.]

CONFIDENCE INTERVAL

A random interval produced by a statistical method that contains the true (fixed, but unknown) parameter value with a stated confidence level on repeated application of the statistical method.

[NOTE—See chapter <1010> for more information.]

CROSSED (AND PARTIALLY CROSSED)

Two factors are crossed (or fully crossed) if each level of each factor appears with each level of the other factor. Two factors are partially crossed when they are not fully crossed but multiple levels of one factor appear with a common level of the other factor.

[NOTE—1. For example, in a bioassay in which all samples appear at all dilutions, samples and dilutions are (fully) crossed. In a bioassay validation experiment in which two of four analysts each perform assays on the same set of samples on each of six days and a different pair of analysts is used on each day, the analysts are partially crossed with days. 2. Each factor may be applied to different experimental units, and the factors may be both fully crossed and nested (q.v.), creating a split-unit or split-plot design (q.v.). 3. Experiments with factors that are partially crossed require particular care for proper analysis. 4. A

randomized complete block design (q.v.) is a design in which the block factor (which often is treated as a random effect) is crossed with the treatment factor (which usually is treated as a fixed effect).]

DESIGN OF EXPERIMENTS (DOE) [ICH Q8(R2)]⁵

A structured, organized method for determining the relationship between factors that affect a process and the output of that process.

[NOTE—DOE is used in bioassay development and validation; see <1032> and <1033>.]

EQUIVALENCE TEST

A test to demonstrate equivalence (e.g., similarity or conformance to validation acceptance criteria) of two quantities by conformance to an interval acceptance criterion.

[NOTE—1. An equivalence test differs from most common statistical tests in the nature of the statistical hypotheses. Most common statistical tests are difference tests—that is, the statistical null hypothesis is that of no difference, and the alternative is that there is some difference, without regard to the magnitude or importance of the difference. The difference may be between a characteristic of two populations or between a characteristic of a single population and an accepted value. In equivalence testing the null hypothesis is that the difference is not sufficiently small, and the alternative hypothesis is that the difference is sufficiently small that there is no important difference. In a common statistical difference test one concludes that there is insufficient evidence to establish nonconformance to an acceptance criterion. This may be the result of excess variability and/or an inadequate design. In an equivalence test the conclusion is that the data conform to the acceptance criterion (e.g., slopes are parallel). 2. A common statistical procedure used for equivalence tests is the two one-sided tests (TOST) procedure. 3. The interval acceptance criterion may be one- or two-sided. An example of a one-sided interval is a validation acceptance criterion for a %GCV of not more than XX%]

EXPECTED MEAN SQUARE

A mathematical expression of variances estimated by an ANOVA mean square.

EXPERIMENTAL DESIGN

The structure of assigning treatments to experimental units.

[NOTE—1. Some aspects of experimental design are blocking (q.v.), randomization (q.v.), replication (q.v.), and specific choice of design (cf. <1032>). 2. Important components of experimental design include the number of samples, the number of concentrations, and how samples and concentrations are assigned to experimental units and are grouped into blocks. 3. The experimental design influences which statistical methodology should be used to achieve the analytical objective.]

EXPERIMENTAL UNIT

The smallest unit to which a distinct level of a treatment is randomly allocated.

[NOTE—1. Randomization of treatment factors to experimental units is essential in bioassays. 2. Different treatment factors can be applied to different experimental units. For example, samples may be assigned to rows on a 96-well plate, and dilutions may be assigned to columns on the plate. In this case, rows are the experimental units for samples, columns are the experimental units for concentrations, and wells are the experimental units for the interaction of sample and concentration. 3. An experimental unit must be distinguished from a sampling unit, the smallest unit on which a distinct measurement is recorded (e.g., a well). Because the sampling unit is often smaller than the experimental unit, it is an easy mistake to treat sampling units as if they are experimental units. This mistake is called pseudoreplication (q.v.).]

FACTOR

An assay parameter or operational element that may affect assay response and that varies either within or across assay runs.

[NOTE—In a bioassay there are at least two treatment factors: sample and concentration.

A **fixed factor (fixed effect)** is a factor that is controllable and deliberately set at specific levels in a bioassay. Inference is made to the levels used in the experiment or intermediate values. Sample and concentration are examples of fixed factors in bioassays.

A **random factor (random effect)** is one which is generally not controllable and for which its levels represent a sample of ways in which that factor might vary. In a bioassay, the test organisms, plate, and day are often considered random factors. Whether a factor is treated as random or fixed may depend on the experiment and questions asked.]

FACTORIAL DESIGN

An experimental design in which there are multiple factors and the factors are partially or fully crossed.

In a **full factorial design**, each level of a factor appears with all combinations of levels of all other factors. For example, if factors are reagent batch and incubation time, for a full factorial design all combinations of incubation time and reagent batch must be included.

⁵ ICH. Guidance Q8(R2) Pharmaceutical Development. November 2009. Available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073507.pdf>. Accessed 27 December 2011.

A **fractional factorial design** is a reduced design in which each level of a factor appears with only a subset of combinations of levels of all other factors and some factor effects (main effects and/or interactions) are deliberately confounded with other combinations of factor effects. Fractional factorial designs should be carefully considered for screening and optimization purposes. This design can be considered without risk of information loss for validation.

GENERAL LINEAR MODEL

A statistical linear model that relates study factors, which can be continuous or discrete, to experimental responses.

INDEPENDENCE

For two measurements or observations *A* and *B* (raw data, assay sets, or relative potencies) to be independent, values for *A* must be unaffected by *B*'s responses and vice versa.

[NOTE—A consequence of the failure to recognize lack of independence is poor characterization of variance. In practice this means that if two potency or relative potency measurements share a common factor that might influence assay outcome (e.g., analyst, cell preparation, incubator, group of animals, or aliquot of Standard samples), then the correct initial assumption is that these relative potency measurements are not independent. The same concern for lack of independence holds if the two potency or relative potency measurements are estimated together from the same model or are in any way associated without including in the model some term that captures the fact that there are two or more potency measurements. As assay experience is gained, an empirical basis may be established (and monitored) so that it is reasonable to treat potency measurements as independent even if the measurements share a common level of a factor. This is the case when it has been demonstrated that a factor does not have a practically significant effect on long-term bioassay results.]

INTERACTION

Two factors are said to interact if the response to one factor depends on the level of the other factor.

LEVEL

A location on the scale of measurement of a factor.

[NOTE—1. Factors have two or more distinct levels. For example, if a bioassay validation experiment employs three values of incubation time and two batches of a key reagent, the levels are the three times for the factor *incubation time* and the two batches for the factor *batch*. 2. Levels of a factor in a bioassay may be quantitative, such as concentration, or categorical, such as sample (i.e., test and reference).]

LOGNORMAL DISTRIBUTION

A distribution of values (assay responses or potencies) where the logarithms of the values have a normal distribution.

[NOTE—1. Most relative potency bioassay measurements are lognormally distributed. 2. The lognormal is a skewed distribution characterized by increased variability with increased level of response.]

MEAN SQUARE

A calculation in ANOVA representing the variability associated with an experimental factor.

MIXED-EFFECTS MODEL

A statistical model that includes both fixed and random effects.

MODELING, STATISTICAL

The mathematical specification of the relationship between inputs (*X*s) and outputs (*Y*s) of a process, e.g., the concentration–response relationship in bioassay or the modeling of the effects of important sources of variation on potency measurement.

[NOTE—1. Modeling includes methods to capture the dependence of the response on the samples, concentration, experimental units, and groups or blocking factors in the assay configuration. 2. Modeling of bioassay data includes making many choices, some of which are driven by the assay design and data. For continuous data there is a choice between linear and nonlinear models. For discrete data there is a choice among logit/log models within a larger family of generalized linear models. In limiting dilution assays, published literature advocates Poisson models and Markov chain binomial models. One can use either fixed-effects models or mixed-effects models for bioassay data. On the one hand, the fixed-effects models are more widely available in software and are somewhat less demanding for statisticians to set up. On the other hand, mixed models have advantages over fixed ones: they are more accommodating of missing data and, more importantly, can allow each block to have different slopes, asymptotes, median effective concentrations required to induce a 50% effect (EC_{50}), or relative potencies. Particularly when the analyst is using straight-line models fitted to nonlinear responses or assay systems in which the concentration–response curve varies from block to block, the mixed model captures the behavior of the assay system in a much more realistic and interpretable way. 3. It is essential that any modeling approach for bioassay data should use all available data simultaneously to estimate the variation (or, in a mixed model, each of several sources of variation). It may be necessary to transform the observations before this modeling to include a variance model or to fit a means model (in which there is a predicted effect for each combination of sample and concentration) to get pooled estimate(s) of variation.]

NESTED

A factor *A* is nested within another factor *B* if the levels of *A* are different for every level of *B*.

[NOTE—1. For example, in a bioassay validation experiment two analysts may perform assays on five days each. If the calendar days for the first analyst are distinct from those of the second analyst, days are nested within analyst. 2. Nested factors have a hierarchical relationship. 3. For two factors to be nested they must satisfy the following: (a) they are applied to different-sized experimental units; (b) the larger experimental unit contains more than one of the smaller experimental units; and (c) the factor applied to the smaller experimental unit is not fully crossed (q.v.) with the factor applied to the larger experimental unit. When conditions (a) and (b) are satisfied and the factors are partially crossed, then the experiment is partially crossed and partially nested. Experiments with this structure require particular care for proper analysis.]

PARALLELISM (OF CONCENTRATION–RESPONSE CURVES)

A quality in which the concentration–response curves of the Test sample and the Reference Standard are identical in shape and differ only by a horizontal difference that is a constant function of relative potency.

[NOTE—1. When Test and Reference preparations are similar (q.v.) and assay responses are plotted against log concentrations, the resulting curve for the Test preparation will be the same as that for the Standard but will be shifted horizontally by an amount that is the logarithm of the relative potency. Because of this relationship, similarity (q.v.) is often equated with *parallelism* but they are not the same. See section 3.5, *Slope-Ratio Concentration–Response Models*, in chapter <1034>, in which similar samples have concentration–response relationships with a common (or nearly common) *y*-intercept but may differ in their slopes. 2. In practice, it is not possible to demonstrate that the shapes of two curves are identical. Instead, the two curves are shown to be sufficiently algebraically similar (equivalent) in shape. Note that *similar* should be interpreted as “we have evidence that the two curves are close enough in shape” rather than “we do not have evidence that the two curves differ in shape.” 3. The assessment of parallelism depends on the type of function used to fit the response curve. Parallelism for a nonlinear assay using a four-parameter logistic fit means that (a) the slopes of the rapidly changing parts of the Test and Reference Standard curves (that is, slope at a tangent to the curve where the first derivative is at a maximum) should be similar; and (b) the upper and lower asymptotes of the response curves (plateaus) should be similar. For straight-line analysis, the slopes of the lines should be similar.]

POINT ESTIMATE

A single-value estimate obtained from statistical calculations.

[NOTE—1. Examples are the average relative bias, the %GCV, and relative potency. 2. The point estimate may be augmented with an *interval estimate* (confidence interval; q.v.) that employs an interval to express the uncertainty in the determination of the point estimate.]

PSEUDOREPLICATION

The misidentification of samples from experimental units as independent and thus true replicates when they actually are not independent.

[NOTE—1. Pseudoreplication results in incorrect inferences because of the incorrect assignment of variability and the appearance of more replicates than are actually present. 2. Lack of recognition of pseudoreplication is critical because it is an easy mistake to make, and the consequences can be serious. For example, pseudoreplicates commonly arise when analysts make a dilution series for each sample in tubes (the dilution series can be made with serial dilutions, single-point dilutions, or any convenient dilution scheme). The analyst then transfers each dilution of each sample to several wells on one or more assay plates. The wells are then pseudoreplicates because they are simply aliquots of a single dilution process and thus are not representative of independent preparations. 3. A simple way to analyze data from pseudoreplicates is to average over the pseudoreplicates (if a transformation of the observed data is used, the transformation should be applied before averaging over pseudoreplicates) before fitting any concentration–response model. In many assay systems, averaging over pseudoreplicates leaves the assay without any replication. A more complex way to use data containing pseudoreplicates is to use a mixed model that treats the pseudoreplicates as a separate random effect. Although pseudoreplication normally is of little value, it can be advantageous when two conditions are satisfied: (a) the pseudoreplicate (e.g., well-to-well) variation is very large compared to the variation associated with replicates; and (b) the cost of pseudoreplicates is much lower than the cost of replicate experimental units.]

P VALUE (SIGNIFICANCE PROBABILITY)

The probability of observing, in repeated trials, that an experimental outcome is as different or more different than that observed if the null hypothesis is true.

[NOTE—1. *More different* means further from the null hypothesis. 2. Commonly, $P < 0.05$ is taken as a threshold for indicating statistically significant differences, although any value for the threshold may be used. Bases for choosing the threshold are the risks (costs) of making a wrong decision; see *type I error* and *type II error*.]

RANDOMIZATION

A process of assignment of treatment to experimental units based on chance so that all equal-sized subgroups of units have an equal chance of receiving a given treatment.

[NOTE—1. The chance mechanism may be an unbiased physical process (rolling unbiased dice, flipping coins, drawing from a well-mixed urn), random-number tables, or computer-generated randomized numbers. Care must be taken in the choice and

use of method. Good practice is to use a validated computerized random-number generator. 2. The use of randomization can help to prevent systematic error from becoming associated with particular samples or a dilution pattern and causing bias. For example, in 96-well bioassays, plate effects can be substantial and can cause bias in observed responses or summary measures. In animal studies, a variety of factors associated with individual animals can influence responses. If extraneous factors that influence either plate assays or animal assays are not routinely demonstrated to have been eliminated or minimized so as to be negligible, randomization is essential to obtaining unbiased data required for the calculation of true potency. 3. Randomization is a good practice even when there is evidence that operational factors (e.g., location, time, reagent lot) have little or no effect on the assay system. While randomization may not protect an individual assay (or perhaps a block of an assay) from a (perhaps newly) important operational factor, randomization provides assurance that results from a collection of assays are not biased due to operational factors.]

REPLICATION

A process in which multiple independent experimental units receive the same level of a treatment factor.

[NOTE—1. The purpose of replication is to minimize the effects of uncontrollable sources of random variability. 2. Replication can occur either completely at random or across blocks. Generally, replication within blocks is pseudoreplication (q.v.). 3. Replication of factors that contribute most greatly to variability, or factors that are at the highest levels in a nested layout, usually result in the most effective reduction of random variability.]

TRUE REPLICATES

Samples based on independent experimental units.

STANDARD ERROR OF ESTIMATE

A measure of uncertainty of an estimate of a reportable value or other parameter estimate because of sampling variation.

[NOTE—1. In bioassay the focus is on the precision (standard error) of the relative potency. 2. Standard errors can be made smaller with additional replication. 3. Technically, the standard error of an estimate is the standard deviation of the sampling distribution of the estimate. The term *standard error* is used to distinguish between this usage of standard deviation (that depends on sample size) and the common laboratory usage in which standard deviation (or coefficient of variation) is used to characterize the precision of individual measurements obtained from a procedure. This latter precision does not depend on sample size.]

STATISTICAL PROCESS CONTROL

A set of statistical methods used to monitor shifts and trends in a process.

TYPE I ERROR

The error in statistical hypothesis testing that the alternative hypothesis is accepted when it is false.

[NOTE—The probability of a type I error usually is denoted by α .]

TYPE II ERROR

The error in statistical hypothesis testing that the alternative hypothesis is rejected when it is true.

[NOTE—The probability of a type II error usually is denoted by β .]

VARIANCE COMPONENT ANALYSIS

A statistical analysis that partitions contributions made to total variability by components associated with influential assay factors, e.g., analyst, day, or instrument.