

DESIGN AND DEVELOPMENT OF BIOLOGICAL ASSAY AND ITS VALIDATION

1. INTRODUCTION

1.1 Purpose and Scope

The purpose of the chapter Development of Biological Assays and its Validation presents methodology for the development of bioassay procedures that have sound experimental design, that provide data that can be analyzed using well-founded statistical principles, and its validation that are fit for their specific use. This general chapter is intended to guide the development of a bioassay for a drug substance or product intended for commercial distribution. Although adoption of this chapter's recommended methods may be resource intensive during assay development, early implementation can yield benefits.

This chapter is intended for both the practicing bioassay analyst who is engaged in biological assay Development, Validation and Analysis of bioassay, later on the assay uses for lot release of drug Substance, Drug Product, In-process sample and its stability testing. It is a good strategy to develop and validate a bioassay to support multiple intended uses; for example, a bioassay primarily developed for batch release may serve other purposes. It is generally useful to assess similarity using the dose response curve, range of linear point on the curve, maximum response etc., The former will find guidance for implementing bioassay structure and methodology to achieve analytical goals while reliably demonstrating the biological activity of interest, and the latter will gain insights regarding the constraints of biology and variability of assay that can prove challenging to balance with a help of statistical tools.

1.2 Biological Assay/ Potency Assay

Biological assays (also called bioassays) are an integral part of the quality assessment required for the manufacturing and marketing of many biological and some non-biological drug products. Bioassays commonly used for drug potency estimation can be distinguished from chemical tests by their reliance on a biological substrate (e.g., animals, living cells, or functional complexes of target receptors). Because of multiple operational and biological factors arising from this reliance on biology, they typically exhibit a greater variability than do chemically-based tests. Bioassays are one of several physicochemical and biologic tests

Draft for stakeholder review

with procedures and acceptance criteria that control critical quality attributes of a biological drug product.

Since the inherent variability in biological test systems (including that from animals, cells, instruments, reagents, and day-to-day and between-lab), an absolute measure of potency/activity is more variable than a measure of activity relative to a Standard. This has led to the adoption of the relative potency methodology. Assuming that the Standard and Test materials are biologically similar, statistical similarity (a consequence of the Test and Standard similarity) should be present, and the Test sample can be expected to behave like a concentration or dilution of the Standard. Relative potency is a unit less measure obtained from a comparison of the dose-response relationships of Test and Standard drug preparations. For the purpose of the relative comparison of Test to Standard, the potency of the Standard is usually assigned a value of 1 (or 100%). The Standard can be a material established as such by a national (e.g., USP) or international (e.g., WHO) organization, or it could be an internal Standard.

2.0 DEVELOPMENT OF BIOLOGICAL ASSAY

2.1 USE OF BIOASSAY

2.1.1 In-process and process Intermediate

Bioassays are generally required in the development and optimization of product manufacturing, including formulation and scale-up processes. Bioassays can be used to evaluate purification strategies, optimize product yield, and measure product stability. The bioassay's precision and accuracy should be sufficient for measuring process performance or for assessing and comparing the stability of candidate formulations.

2.1.2 Product Release

Bioassays are used to evaluate the potency of the drug before commercial product release. To the extent possible, the assay should reflect or mimic the product's known or intended mechanism of action. If the bioassay does not include the functional biology directly associated with the mechanism of action, it may be necessary to demonstrate a relationship between the bioassay's estimated potency determinations and those of some other assay that better or otherwise reflects putative functional activity. For product-release testing, product specifications are established to define a minimum or range of potency values that are

Draft for stakeholder review

acceptable for product. The precision of the reportable value from the bioassay must support the number of significant digits listed in the specification.

2.1.3 Stability

The potency assay may be used to assess biotechnology and vaccine product stability. Information from stability studies, performed during development under actual and/or accelerated or stressed storage conditions, may be used to establish shelf life duration as well as to identify and estimate degradation products and degradation rates. Post licensure stability studies may be used to monitor product stability. Knowledge of both short-term and long-term variability of the bioassay is important to assure an acceptable level of uncertainty in potency measures obtained.

2.2 Basics of Biological assay

2.2.1 *In-Vitro* (Cell-Based) Bioassays

Bioassays using cell lines that respond to specific ligands or infectious agents can be used for lot-release assays. These cell lines can be derived from tumors, immortalized as factor-dependent cell lines, or engineered cell lines transfected with appropriate receptors. Additionally, non-transformed cell lines which can be maintained over a sufficient number of passages (e.g., fibroblasts) may also be used. Regardless of cell line, there is an expectation of adequately equivalent potency response through some number of continuous passages to be set. Advances in recombinant DNA technology and the understanding of cellular signaling mechanisms have allowed the generation of engineered cell lines with improved response, stable expression of receptor and signalling mechanisms, and longer stability. The cellular responses to the protein of interest depend on the drug's MOA (Mechanism OF Action) and the duration of exposure. Such responses include cell proliferation, cell killing, antiviral activity, differentiation, cytokine/mediator secretion, and enzyme activation. Assays involving these responses may require incubation of the cells over several days, during which time contamination, uneven evaporation, or other location effects may arise. Comparatively rapid responses based on an intracellular signalling mechanism—such as second messengers assay, protein kinase activation, or reporter gene expression—have proven acceptable to regulatory authorities. Lastly, most cell lines used for bioassays express receptors for multiple cytokines and growth factors. This lack of specificity may not be detrimental if the Test sample's specificity is demonstrated.

Draft for stakeholder review

Cell-based bioassay design should reflect knowledge of the factors that influence the response of the cells to the active analyte. Response variability is often reflected in parameters such as slope, EC_{50} of the concentration–response curve, or the response range (maximum minus minimum response). Even though relative potency methodology minimizes the effects on potency estimates of variation in these parameters among assays, and among blocks within an assay, such response variability can make an assay difficult to manage (i.e., it may be difficult to assess system suitability). Hence, while assay development should be focused primarily on the properties of potency, efforts to identify and control variation in the concentration–response relationship are also appropriate. For blocked assays (e.g., multiple cell culture plates in an assay) with appreciable variation in curve shape among blocks, an analysis that does not properly include blocks will yield inflated estimates of within assay variation, making similarity assessment particularly difficult. Two strategies are available for addressing variation among blocks: one, a laboratory effort to identify and control sources of variation and two, a statistical effort to build and use a blocked design and analysis. Combining these strategies can be particularly effective.

The development of a cell-based bioassay begins with the selection or generation of a cell line. An important first step when developing a cell-based assay to assess a commercial product is to verify that the cell line of interest is not restricted to research use only. To ensure an adequate and consistent supply of cells for product testing, a cell bank should be generated if possible. To the extent possible, information regarding functional and genetic characteristics of the bioassay's cell line should be documented, including details of the cell line's history from origin to banking. For example, for a recombinant cell line this might include the identification of the source of the parental cell line (internal cell bank, external repository, etc.), of the DNA sequences used for transfection, and of the subsequent selection and functional testing regimen that resulted in selection of the cell line. Ideally, though not always practical, sufficient information is available to permit recreation of a similar cell line if necessary. Pertinent information may include

- a. Identity (e.g., isoenzyme, phenotypic markers, genetic analysis);
- b. Morphology (e.g., archived photographic images);
- c. Purity (e.g., mycoplasma, bacteria, fungus and virus testing);

Draft for stakeholder review

- d. Cryopreservation; thaw and culture conditions (e.g., media components, thaw temperature and method, methods of propagation, seeding densities, harvest conditions);
- e. Thaw viability (immediately after being frozen and after time in storage);
- f. Growth characteristics (e.g., cell doubling times); and
- g. Functional stability (e.g., ploidy).

Cell characterization and vigilance regarding aspects of assay performance that reflect on cell status are necessary to ensure the quality and longevity of cell banks for use in the QC environment.

The general health and metabolic state of the cells at the time of bioassay can substantially influence the test results. After a cell line has been characterized and is ready for banking, analysts typically prepare a two-tiered bank (Master and Working). A Master Cell Bank is created as the source for the Working Cell Bank. The Working Cell Bank is derived by expansion of one or more vials of the Master Cell Bank. The size of the banks depends on the growth characteristics of the cells, the number of cells required for each assay, and how often the assay will be performed. Some cells may be sensitive to cryopreservation, thawing, and culture conditions and the banks must be carefully prepared and characterized before being used for validation studies and for regular use in the QC laboratory.

There follow factors that may affect bioassay response and the assessment of potency, that are common to many cell-based bioassays:

- a. Cell type (adherent or non-adherent);
- b. Cell thawing; plating density (at thaw and during seed train maintenance) and confluence (adherent cells);
- c. Culture vessels; growth, staging, and assay media; serum requirements (source, heat inactivation, gamma irradiation);
- d. Incubation conditions (temperature, CO₂, humidity, culture times from thaw);
- e. Cell harvesting reagents and techniques (for adherent cells, method of dissociation);
- f. Cell sorting & cell counting;
- g. Determination of cell health (growth rate, viability, yield);
- h. Cell passage number and passaging schedule;
- i. Cell line stability (genetic, receptor, marker, gene expression level); and
- j. Starvation or stimulation steps.

Draft for stakeholder review

This list is not exhaustive, and analysts with comprehensive understanding and experience with the cell line should be involved during assay development. These experienced individuals should identify factors that might influence assay outcomes and establish strategies for an appropriate level of control whenever possible.

2.2.2 *In-Vivo* Bioassays

In vivo potency assays are bioassays in which sets of dilutions of the Standard and Test materials are administered to animals and the concentration–response relationships are used to estimate potency. For some animal assays, the endpoint is simple (e.g., rat body weight gain assay for human growth hormone or rat ovarian weight assay for follicle stimulating hormone), but others require further processing of samples collected from treated animals (e.g., reticulocyte count for erythropoietin, steroidogenesis for gonadotropins, neutrophil count for granulocyte colony stimulating factor, or antibody titer after administration of vaccines). With the advent of cell lines specific for the putative physiological mechanism of action (MOA), the use of animals for the measurement of potency has substantially diminished. Cost, low throughput, ethical, and other practical issues argue against the use of animal bioassays. Selection of animal test subjects by species, strain, gender, and maturity or weight range is guided by the goal of developing a representative and sensitive model with which to assess the activity of Test samples. Some assay methods lend themselves to the use of colony versus naive animals. For example, pyrogen and insulin testing benefit from using experienced colony rabbits that provide a reliable response capacity. If animals recently introduced to the colony fail to respond as expected after several administrations of a compound, they should be culled from the colony so they do not cause future invalid or indeterminate assay results. In the case of assaying highly antigenic compounds for pyrogens, however, naive animals should be used to avoid generating inaccurate or confounded results. Other colony advantages include common controlled environmental conditions (macro/room, and micro/rack), consistent feeding schedule, provision of water, and husbandry routine. Historical data including colony records and assay data can be used to identify factors that influence assay performance. The influence of biasing factors can be reduced by applying randomization principles such as distribution of weight ranges across dose groups, group assignments from shipping containers to different cages, or use of computer-generated or deck patterns for injection/dosing. A test animal must be healthy and have time to stabilize in its environment to be suitable for use in a bioassay. Factors that combine to influence an

Draft for stakeholder review

animal's state of health include proper nutrition, hydration, freedom from physical and psychological stressors, adequate housing sanitization, controlled light cycle (diurnal/nocturnal), experienced handling, skillful injections and bleedings, and absence of noise or vibration. Daily observation of test animals is essential for maintenance of health, and veterinary care must be available to evaluate issues that have the potential to compromise the validity of bioassay results.

2.2.3 *Ex-Vivo* Bioassays

Cells or tissues from human or animal donors can be cultured in the laboratory and used to assess the activity of a Test sample. In the case of cytokines, the majority of assays use cells from the hematopoietic system or subsets of hematopoietic cells from peripheral blood such as peripheral blood mononuclear cells or peripheral blood lymphocytes. For proteins that act on solid tissues, such as growth factors and hormones, specific tissue on which they act can be removed from animals, dissociated, and cultured for a limited period either as adherent or semi-adherent cells. Although an *ex vivo* assay system has the advantage of similarity to the natural milieu, it may also suffer from substantial donor-to-donor variability, as well as challenging availability of appropriate cells. Bioassays that involve live tissues or cells from an animal (e.g., rat hepatocyte glucagon method) require process management similar to that of *in vivo* assays to minimize assay variability and bias. The level of effort to manage bias (e.g., via randomization) should be appropriate for the purpose of the assay. Additional factors that may affect assay results include time of day, weight or maturity of animal, anesthetic used, buffer components/reagents, incubation bath temperature and position, and cell viability.

2.2.4 Reagent qualification

The quantitative characterization of a new Standard requires an accurate and precise measurement of the new Standard's biological activity. This measurement is used either to establish that the new Standard lot is equivalent to the previous lot or to assign it a label potency to which Test samples can be compared. Additional replication (beyond routine testing) may be required to achieve greater precision in the potency measurement of the new Standard material. Additionally, the bioassay may be used to qualify a cell culture reagent such as fetal bovine serum. The fitness for use in such cases is tied to the ability of the assay

Draft for stakeholder review

to screen reagent lots and to ensure that lots that may bias or compromise the relative potency measurements are not accepted.

2.2.5 Reference Standard

The Standard is a critical reagent in bioassays because of the necessity to have a reliable material to which a Test preparation can be quantitatively compared. The Standard may be assigned a unitage or specific activity that represents fully (100%) potent material. Where possible, a Standard should be representative of the samples to be tested in the bioassay. Testing performed to qualify a Standard may be more rigorous than the routine testing used for lot release. A Standard must be stored under conditions that preserve its full potency for the intended duration of its use. To this end, the Standard may be stored under conditions that are different from the normal storage of the drug substance or drug product. These could include a different temperature (e.g., -70° or -20° , instead of 2° – 8°), a different container (e.g., plastic vials instead of syringes), a different formulation (e.g., lyophilizable formulation or the addition of carrier proteins such as human serum albumin, stabilizers, etc.). The Standard material should be tested for stability at appropriate intervals.

System suitability criteria of the bioassay such as maximum or background response, EC_{50} slope, or potency of assay control may be used to detect change in the activity of the Standard. Accelerated stability studies can be performed to estimate degradation rates and establish recognizable characteristics of Standard instability. At later stages in clinical development, the Standard may be prepared using the manufacturing process employed in pivotal clinical trials. If the Standard formulation is different from that used in the drug product process, it is important to demonstrate that the assay's assessment of similarity and estimate of potency is not sensitive to the differences in formulation. An initial Standard may be referred to as the Primary Standard. Subsequent Standards can be prepared using current manufacturing processes and can be designated Working Standards. Separate SOPs may be required for establishing these standards for each product. Bias in potency measurements sometimes can arise if the activity of the Standard gradually changes. Also, loss of similarity may be observed if, with time, the Standard undergoes changes in glycosylation. It is prudent to archive aliquots of each Standard lot for assessment of comparability with later Standards and for the investigation of assay drift.

2.3 STATISTICAL ASPECTS OF BIOASSAYS

Draft for stakeholder review

The statistical elements of bioassay development include the type of data, the measure of response at varying concentration, the assay design, the statistical model, pre-analysis treatment of the data, methods of data analysis, suitability testing, and outlier analysis. These form the constituents of the bioassay system that will be used to estimate the potency of a Test sample.

2.3.1 Data

Fundamentally, there are two bioassay data types: quantitative and quantal (categorical). Quantitative data can be either continuous (not limited to discrete observations; e.g., collected from an instrument), count (e.g., plaque-forming units), or discrete (e.g., endpoint dilution titers). Quantal data are often dichotomous; for example, life/death in an animal response model or positivity/negativity in a plate-based infectivity assays those results in destruction of a cell monolayer following administration of an infectious agent. Quantitative data can be transformed to quantal data by selecting a threshold that distinguishes a positive response from a negative response. Such a threshold can be calculated from data acquired from a negative control, as by adding (or subtracting) a measure of uncertainty (such as two or three times the standard deviation of negative control responses) to the negative control average. Analysts should be cautious about transforming quantitative data to quantal data because this results in a loss of information.

2.3.2 Assumptions for data similarity

A key assumption for the analysis of most bioassays is that the Standard and Test samples contain the same effective analyte or population of analytes and thus may be expected to behave similarly in the bioassay. This is termed similarity for specific statistical models, biological similarity implies that statistical similarity is present (for parallel-line and parallel-curve models, the Standard and Test curves are parallel; for slope-ratio models, the Standard and Test lines have a common intercept). The reverse is not true. Statistical similarity (parallel lines, parallel curves, or common intercept, as appropriate) does not ensure biological similarity. However, failure to satisfy statistical similarity may be taken as evidence against biological similarity. The existence of a Standard–Test sample pair that passes the assessment of statistical similarity is thus a necessary but not sufficient condition for the satisfaction of the key assumption of biological similarity. Biological similarity thus remains, unavoidably, an assumption. Departures from statistical similarity that are consistent in value across replicate assays may be indicative of matrix effects or of real differences between Test and

Draft for stakeholder review

Standard materials. This is true even if the departure from statistical similarity is sufficiently small to support determination of a relative potency.

In many assays multiple compounds will yield similar concentration–response curves. It may be reasonable to use a biological assay system to describe or even compare response curves from different compounds. But it is not appropriate to report relative potency unless the Standard and Test samples contain only the same active analyte or population of analytes. Biological products typically exhibit lot-to-lot variation in the distribution of analytes (i.e., most biological products contain an intended product and, at acceptably low levels, some process contaminants that may be active in the bioassay). Assessment of similarity is then, at least partially, an assessment of whether the distribution of analytes in the Test sample is close enough to that of the distribution in the Standard sample for relative potency to be meaningful; that is, the assay is a comparison of like to like. When there is evidence (from methods other than the bioassay) that the Standard and Test samples do not contain the same active compound(s), the assumption of biological similarity is not satisfied, and it is not appropriate to report relative potency.

2.3.3 Assay variance, weighting and transformation

For linear and nonlinear regression models, the variance referred to here is the residual variance from the fit of the model. Constant variance is often not observed; variance heterogeneity may manifest as an increase in variability with increase in response. If the variances are not equal but the data are analyzed as though they are, the estimate of relative potency may still be reasonable; however, failure to address non-constant variance around the fitted concentration–response model results in an unreliable estimate of within-assay variance. Further, the assessment of statistical similarity may not be accurate, and standard errors and confidence intervals for all parameters (including a Fieller’s Theorem-based interval for the relative potency) should not be used. Confidence intervals for relative potency that combine potency estimates from multiple assays may be erroneous if within-assay error is used for confidence interval calculation.

Variance is best assessed on a large body of assay data. Using only the variance among replicates from the current assay is not appropriate, because there are too few data to properly determine truly representative variances specific to each concentration. Data on variance is

Draft for stakeholder review

sparse during development; it is prudent to re-assess variance during validation and to monitor it periodically during ongoing use of the assay. Two methods used to mitigate variance heterogeneity are transformation and weighting. Lack of constant variance can be addressed with a suitable transformation. Additionally, transformation can improve the normality of residuals and the fit of some statistical models to the data. A transformation should be chosen for an assay system during development, checked during validation, used consistently in routine assay practice, and checked periodically.

Bioassay data are commonly displayed with log-transformed concentration; slope-ratio assays are displayed with concentration on the original scale. Transformation may be performed to the response data as well as to the concentration data. Common choices for a transformation of the response include log, square root (for counts), reciprocal, and, for count data with known asymptotes, logit of the percent of maximum response. Log transformations are commonly used, as they may make nearly linear a useful segment of the concentration–response relationship, and because of the ease of transforming back to the original scale for interpretation. A log–log fit may be performed on data exhibiting nonlinear behavior. Other alternatives are available; i.e., data may be transformed by the inverse of the Power of the Mean (POM) function. A POM coefficient of $k = 2$ corresponds to a log transformation of the data.

Note that transformation of the data requires re-evaluation of the model used to fit the data. From a statistical perspective there is nothing special about the original scale of measurement; any transformation that improves accordance with assumptions is acceptable. Analysts should recognize, however, that transformations, choice of statistical model, and choice of weighting scheme are interrelated. If a transformation is used, that may affect the choice of model. That is, transforming the response by a log or square root, for example, may change the shape of the response curve, and, for a linear model, may change the range of concentrations for which the responses are nearly straight and nearly parallel.

For assays with non-constant variance, a weighted analysis may be a reasonable option. Though weighting cannot address lack of residual normality, it is a valid statistical approach to placing emphasis on more precise data. Ideally, weights may be based on the inverse of the predicted within-assay (or within-block) variance of each response where the predictors of

Draft for stakeholder review

variance are independent of responses observed in a specific assay. In practice, many bioassays have relatively large variation in $\log EC_{50}$ (compared to the variation in \log relative potency) among assays (and sometimes among blocks within assay). If not addressed in the variance model, this variation in $\log EC_{50}$ induces what appears to be large variation in response near the mean $\log EC_{50}$, often yielding too-low weights for observations near the EC_{50} .

If the assay is fairly stable (low variability in EC_{50}), an alternative is to look at variance as a function of concentration. While not ideal, an approach using concentration-dependent variances may be reasonable when the weights are estimated from a large number of assays, the variances are small, any imbalance in the number of observations across concentrations is addressed in the variance model, and there are no unusual observations (outliers). This possibility can be examined by plotting the response variance at each concentration (preferably pooled across multiple assays) against concentration and then against a function of concentration (e.g., concentration squared). Variance will be proportional to the function of concentration where this plot approximates a straight line. The apparent slope of this line is informative, in that a horizontal line indicates no weighting is needed. If a function that yields a linear plot can be found, then the weights are taken as proportional to the reciprocal of that function. There may be no such function, particularly if the variation is higher (or lower) at both extremes of the concentration range studied.

Whether a model or historical data are used, the goal is to capture the relative variability at each concentration. It is not necessary to assume that the absolute level of variability of the current assay is identical to that of the data used to determine the weighting, but only that the ratios of variances among concentrations are consistent with the historical data or the data used to determine the variance function.

Appropriate training and experience in statistical methods are essential in determining an appropriate variance-modelling strategy. Sources of variability may be misidentified if the wrong variance model is used. For example, data may have constant variation throughout a four-parameter logistic concentration–response curve but can also have appreciable variation in the EC_{50} parameter from block to block within the assay, or from assay to assay. If the between-block or between-assay variability is not recognized, this assay can appear to have

Draft for stakeholder review

large variation in the response for concentrations near the long-term average value of the EC_{50} . A weighted model with low weights for concentrations near the EC_{50} would misrepresent a major feature of such an assay system.

2.3.4 Linearity of Concentration–Response

The bioassay data responses assume that the shape of the concentration–response curve is a straight line or approximates a straight line over a limited range of concentrations. In those cases, a linear-response model may be assessed to determine if it is justified for the data in hand. Difference testing methods for assessing linearity face the same problems as do difference testing methods applied to parallelism—more data and better precision make it more likely to detect nonlinearity in such case the significant level should be evaluated (95% to 99.99%) during development of the assay and shall be validated during validation of bioassay. Because instances in which lack of linearity does not affect the potency estimate are rare, analysts should routinely assess departure from linearity if they wish to use a linear-response model to estimate potency. If an examination of a data plot clearly reveals departure from linearity, this is sufficient to support a conclusion that linearity is not present. High data variability, however, may mask departure from linearity.

Often a subset of the concentrations measured in the assay will be selected in order to establish a linear concentration–response curve. The subset may be identified graphically. The concentrations at the extreme ends of the range should be examined carefully as these often have a large impact on the slope and calculations derived from the slope. If, in the final assay, the intent is to use only concentrations in the linear range, choose a range of concentrations that will yield parallel straight lines for the relative potencies expected during routine use of the assay; otherwise, the assay will fail parallelism tests when the potency produces assay response values outside the linear range of response. When potency is outside the linear range, it may be appropriate to adjust the sample concentration based on this estimated potency and test again in order to obtain a valid potency result.

The repeat assays together with the valid assays may generate a biased estimate of potency because of the selective process of repeating assays when the response is in the extremes of the concentration–response curve. The problem is more complex in assays where there is even modest variation in the shape or location of the concentration–response curve from run to run or from block to block within an assay. In such assays it may be appropriate to choose subsets

Draft for stakeholder review

for each sample in each assay or even in each block within an assay. Note that a fixed-effects model will mask any need for different subsets in different blocks, but a mixed-effects model may reveal and accommodate different subsets in different blocks (see section Fixed and Random Effects in Models of Bioassay Response).

Additional guidance about selection of data subset(s) for linear model estimation of relative potency includes the following: use at least three, and preferably four, adjacent concentrations; require that the slope of the linear segment is sufficiently steep; require that the lines fit to Standard and Test samples are straight; and require that the fit regression lines are parallel. One way to derive a steepness criterion is to compute a t-statistic on the slope difference from zero. If the slope is not significant the bioassay is likely to have poor performance; this may be observed as increased variation in the potency results. Another aspect that supports requiring adequate steepness of slope is the use of subset selection algorithms. Without a slope steepness criterion, a subset selection algorithm that seeks to identify subsets of three or more contiguous data points that are straight and parallel might select concentrations on an asymptote. Such subsets are obviously inappropriate to use for potency estimation. How steep or how significant the steepness of the slope should be depends on the assay. This criterion should be set during assay development and possibly refined during assay validation.

2.3.5 Common Bioassay Models

Most bioassays consist of a series of concentrations or dilutions of both a Test sample and a Standard material. A mathematical model is fit to the concentration–response data, and a relative potency may then be calculated from the parameters of the model. Choice of model may depend on whether quantitative or qualitative data are being analyzed. For quantitative data, models using parallel response profiles which support comparative evaluation for determining relative potency may provide statistical advantages. If such a model is used, concentrations or dilutions are usually scaled geometrically, e.g., usually in two-fold, log, or half-log increments. If a slope-ratio model is used, concentrations or dilutions can be equally spaced on concentration, rather than log concentration. Several functions may be used for fitting a parallel response model to quantitative data, including a linear function, a higher-order polynomial function, a four-parameter logistic (symmetric sigmoid) function, and a

Draft for stakeholder review

five-parameter logistic function for asymmetric sigmoids. Such functions require a sufficient number of concentrations or dilutions to fit the model.

To assess lack of fit of any model it is necessary to have at least one, and preferably several, more concentrations (or dilutions) than the number of parameters that will be estimated in the model. Also, at least one, and better, two, concentrations are commonly used to support each asymptote. A linear model is sometimes selected because of apparent efficiency and ease of processing. ***Because bioassay response profiles are usually nonlinear, the laboratory might perform an experiment with a wide range of concentrations in order to identify the approximately linear region of the concentration–response profile.*** For data that follow a four-parameter logistic model, these are the concentrations near the centre of the response region, often from 20% to 80% response when the data are rescaled to the asymptotes. Caution is appropriate in using a linear model because for a variety of reasons the apparently linear region may shift. A stable linear region may be identified after sufficient experience with the assay and with the variety of samples that are expected to be tested in the assay. Data following the four-parameter logistic function may also be linearized by transformation. The lower region of the function is approximately linear when the data are log transformed (log–log fit).

2.3.6 Assay Suitability Testing

System Suitability and sample suitability assessment should be performed to ensure the quality of bioassay results. System suitability in bioassay, as in other analytical methods, consists of pre-specified criteria by which the validity of an assay (or, perhaps, a run containing several assays) is assessed. Analysts may assess system suitability by determining that some of the parameters of the Standard response are in their usual ranges and that some properties (e.g., residual variation) of the data are in their usual range. To achieve high assay acceptance rates, it is advisable to accept large fractions of these usual ranges (99% or more i.e. 99.99% for linearity) and to assess system suitability using only a few uncorrelated Standard response parameters. The choice of system suitability parameters and their ranges may also be informed by empirical or simulation studies that measure the influence of changes in a parameter on potency estimation. Sample suitability in bioassay is evaluated using pre-specified criteria for the validity of the potency estimate of an individual Test sample, and usually focuses on similarity assessment. System and sample suitability criteria

Draft for stakeholder review

should be established during bioassay development and before bioassay validation. Where there is limited experience with the bioassay, these criteria may be considered provisional.

2.3.6.1 System Suitability

System suitability parameters may be selected based on the design and the statistical model. Regardless of the design and model, however, system suitability parameters should be directly related to the quality of the bioassay. These parameters are generally based on standard and control samples. In parallel-line assays, for example, low values of the Standard slope typically yield estimates of potency with low precision. Rather than reject assays with low slope, analysts may find it more effective to use additional replicate assays until the assay system can be improved to consistently yield higher-precision estimates of potency. It may be particularly relevant to monitor the range of response levels and location of asymptotes associated with controls or Standard sample to establish appropriate levels of response. A drift or a trend in some of the criteria may indicate the degradation of a critical reagent or Standard material. Statistical process control (SPC) methods should be implemented to detect trends in system suitability parameters. Two common measures of system suitability are assessment of the adequacy of the model (goodness of fit) and of precision. With replicates in a completely randomized design, a pure error term may be separated from the assessment of lack of fit. Care should be taken in deriving a criterion for lack of fit; the use of the wrong error term may result in an artificial assessment. The lack of fit sum of squares from the model fit to the Standard may, depending on the concentrations used and the way in which the data differ from the model, be a useful measure of model adequacy. A threshold may be established, based on sensitivity analysis (assessment of assay sensitivity to changes in the analyte) and/or historical data, beyond which the lack of fit value indicates that the data are not suitable.

The System Suitability should check at least for parallelism, linearity and regression for linear/ nonlinear models. The significant level for validity test should be evaluated during development and refined during the validation.

Parallelism at 95%

Linearity (lack of fit) at 99.99%

Regression at 95%

2.3.6.2 Sample Suitability

Draft for stakeholder review

Sample suitability in bioassay generally consists of the assessment of similarity, which can only be done within the assay range. Relative potency may be reported only from samples that both show similarity to Standard, exhibit requisite quality of model fit, and have been diluted to yield an EC_{50} (and potency) within the range of the assay system.

2.3.6.3 Similarity

Similarity measures may be based on the parameters of the concentration–response curve and may include the slope for a straight parallel-line assay; intercept for a slope-ratio assay; the slope and asymptotes for a four-parameter logistic parallel-line assay; or the slope, asymptotes, and non-symmetry parameter in a five-parameter sigmoid model. In some cases, these similarity measures have interpretable, practical meaning in the assay; certain changes in curve shape, for example, may be associated with specific changes (e.g., the presence of a specific active contaminant) in the product. When possible, discussion of these changes and their likely effects is a valuable part of setting appropriate equivalence boundaries.

2.3.6.4 Range

The range for a relative potency bioassay is the interval between the upper and lower relative potencies for which the bioassay is shown to have suitable levels of precision, relative accuracy, linearity of log potency, and success rates for system and sample suitability. It is straightforward to determine whether or not a sample that is similar to a Standard has a relative potency within the (validated) range of the assay system. For samples that are not similar according to established criteria, it is more challenging to determine whether a relative potency estimate for the sample might be obtained. In a nonlinear parallel-line assay a sample that does not have data on one asymptote might be assumed to be out of the potency range of the assay. In a parallel straight-line assay a sample that does not have three or more points on the steep portion of the response curve may be out of the potency range of the assay. For samples that have not been shown to be similar to reference it is not appropriate to report potency or to construct a ratio of EC_{50} s from unrestricted fits. As such samples may be out of the assay range, it may be useful to shift the dilution of the test sample for a subsequent assay on the basis of an estimate of relative activity. This estimated relative activity may be obtained via the ratio of the concentrations of Standard and Test that yields responses that match the reference response at the reference EC_{50} .

Draft for stakeholder review

2.4 OUTLIERS

Bioassay data should be screened for outliers before relative potency analysis. Outliers may be simple random events or a signal of a systematic problem in the bioassay. Systematic error that generates outliers may be due to a dilution error at one or more concentrations of a Test sample or the Standard or due to a mechanical error (e.g., system malfunction). Several approaches for outlier detection can be considered. Visual inspection is frequently utilized but should be augmented with a more objective approach to avoid potential bias. An outlier is a datum that appears not to belong among the other data present. An outlier may have a distinct, identifiable cause, such as a mistake in the bench work, equipment malfunction, or a data recording error, or it could just be an unusual value relative to the variability typically seen and may appear without an identifiable cause. The essential question pertaining to an outlier becomes: Is the apparent outlier sampled from the same population as the other, less discordant, data, or is it from another population? If it comes from the same population and the datum is, therefore, an unusual (yet still legitimate) value obtained by chance, then the datum should stand. If it comes from another population and the datum's excursive value is due to human error or instrument malfunction, then the datum should be omitted from calculations.

In practice, the answer to this essential question is often unknown, and investigations into causes are often inconclusive. ***Outlier management relies on procedures and practices to yield the best answer possible to that essential question and to guide response accordingly.*** Outlier analysis techniques appropriate for data obtained from regression of response on concentration can be used; the bioassay may be expected to be the most prone to outlying data. The management of outliers is appropriate with bioassay data on at least two levels: where an individual datum or a group of data (e.g., data at a concentration, including blank) can be checked against expected responses for the sample and concentration; and, separately, when estimates of relative potency from an assay can be checked for consistency with other independent estimates of the potency of the same material. Three important aspects of outlier management are prevention, labelling, and identification. Outlier prevention is preferred for obvious reasons, and is facilitated by procedures that are less subject to error and by checks that are sensitive to the sorts of errors that, given the experience gained in assay development, may be expected to occur.

In effect, the error never becomes an outlier because it is prevented from occurring. Good practice calls for the examination of data for outliers and labelling ("flagging") of the

Draft for stakeholder review

apparently outlying observation(s) for investigation. If investigation finds a cause, then the outlying datum may be excluded from analysis. Because of the ordinary occurrence of substantial variability in bioassay response, a laboratory's investigation into the outlying observation is likely to yield no determinable cause. However, the lack of evidence regarding an outlier's cause is not a clear indication that statistical outlier testing is warranted.

Knowledge of the typical range of assay response variability should be the justification for the use of statistical outlier tests. Outlier identification is the use of rules to confirm that the values are inconsistent with the known or assumed statistical model. For outliers with no determined cause, it is tempting to use statistical outlier identification procedures to discard unusual values. Discarding data solely because of statistical considerations should be a rare event. Falsely discarding data leads to overly optimistic estimates of variability and can bias potency estimates. The laboratory should monitor the failure rate for its outlier procedure and should take action when this is significantly higher than expected. Statistical procedures for outlier identification depend on assumptions about the distribution of the data without outliers. Identification of data as outliers may mean only that the assumption about distribution is not correct. If dropping outliers because of statistical considerations is common, particularly if outliers tend to occur more often at high values or at high responses, then this may be an indication that the data require some adjustment, such as log transformation, as part of the assay procedure. Two approaches to statistical assessment of outlying data are replication-based and model-based.

2.4.1 Replication based approaches

When replicates are performed at concentrations of a Test sample and the Standard, an “extra variability” (EV) criterion may be employed to detect outliers. Historical data can be analyzed to determine the range in variability commonly observed among replicates, and this distribution of ranges can be used to establish an extreme in the range that might signal an outlier. Metrics that can be utilized are the simple range (maximum replicate minus minimum replicate), the standard deviation, or the CV or RSD among replicates. However, if the bioassay exhibits heterogeneity of variability, assumptions about uniform scatter of data are unsupported. Analysts can use a variable criterion across levels in the bioassay, or they can perform a transformation of the data to a scale that yields homogeneity of variability. Transformation can be performed with a POM (Power of Mean) approach. Where heterogeneity exists non-normality is likely present, and the range rather than standard

Draft for stakeholder review

deviation or RSD should be used. The actions taken upon detection of a potential outlier depend in part on the number of replicates. If EV is detected within a pair ($n = 2$) at a concentration of a Test sample or the Standard, it will not always be clear which of the replicates is aberrant, and the laboratory should eliminate the concentration from further processing. If more than two replicates are performed at each dilution the laboratory may choose to adopt a strategy that identifies which of the extremes may be the outlier. Alternatively, the laboratory may choose to eliminate the dilution from further processing.

2.4.2 Model based approaches

Model-based approaches may be used to detect outliers within bioassay data. These approaches use the residuals from the fit of an appropriate model. In general, if using model-based methods to identify potential outliers, the models used may make fewer assumptions about the data than the models used to assess suitability and estimate potency. For example, a non-parametric regression (smoothing) model may be useful. Lastly, an alternative to discarding outlying data is to use robust methods that are less sensitive to influence by outlying observations. Use of the median rather than the mean to describe the data's centre exemplifies a robust perspective. Also, regression using the method of least squares, which underlies many of the methods in this chapter, is not robust in the presence of outliers.

2.5 FIXED AND RANDOM EFFECTS IN MODELS OF BIOASSAY RESPONSE

The choice of treating factors as fixed or random is important for the bioassay design, the development experiments, the statistical analysis of data, and the bioassay validation. Fixed effects are factors for which all levels, or all levels of interest, are discretely present, like sample, concentration, temperature and duration of thaw, and incubation time. Data for a response at some level, or combination of levels, of a fixed factor, can predict future responses. Fixed effects are expected to cause a consistent shift in responses. Analysts study fixed effects by controlling them in the design and examining changes in means across levels of the factor. Random effects are factors of which the levels in a particular run of an assay are considered representative of levels that could be present. That is, there is no expectation that any specific value of the random factor will influence response. Rather, that value may vary subject to some expected distribution of values and thus may be a source of variability. For example, there is no desire to predict assay response for a specific day, but there is interest in predicting the variation in response associated with the factor “day”. Examples of random

Draft for stakeholder review

effects include reagent lot, operator, or day if there is no interest in specific reagent lots, operators, or day as sources of variability. Analysts may study random effects by measuring the variance components corresponding to each random effect. Variance components can be estimated well only if there are an appreciable number of levels of each random effect. If there are, for example, only two or three reagent lots or analysts present, the variation associated with these factors will be poorly estimated. Making a correct choice regarding treating a factor as fixed or random is important to the design of the assay and to proper reporting of its precision. Treating all factors as fixed, for example, leads to an understatement of assay variability because it ignores all sources of variability other than replication. The goal is to identify specific sources of variability that can be controlled, to properly include those factors in the design, and then to include other factors as random. If the factor may switch from random to fixed effect or vice versa, the factor should normally be modelled as a random effect. For example, reagent lots cannot be controlled, so different lots are typically considered to cause variability, and reagent lot would be considered a random effect. However, if a large shift in response values has been traced to a particular lot, a comparison among a set of lots could be performed using reagent lot as a fixed effect. Similarly, within-assay location (e.g., block, plate, plate row, plate column, or well) or sequence may be considered a source of random variation or a source of a consistent (fixed) effect. Assay designs that consist of multiple factors are efficient, but require corresponding statistical techniques that incorporate the factors as fixed or random effects in the analysis. If all factors are fixed, the statistical model is termed a fixed-effects model. If all are random, it is termed a random-effects model. If some factors are fixed and some random, the model is a mixed-effects model. Note that the concepts of fixed and random effects apply to models for quantitative, qualitative and integer responses. For assay designs that include multiple experimental units (e.g., samples assigned to sets of tubes and concentrations assigned to pre-plate tubes) a mixed-effects model in which the experimental units are treated as random effects is particularly effective. Additional complexity is added by the presence of designs with crossed random effects (e.g., each operator used material from one or more reagent batches, but many reagent batches were used by multiple operators). This can cause methodological and computational challenges for model fitting, especially when the designs are unbalanced.

2.6 BIOASSAY DEVELOPMENT PROCESS

Draft for stakeholder review

The development of a cell-based bioassay will be used to illustrate the stages in the bioassay development continuum.

An objective of bioassay development is to achieve optimal bioassay relative accuracy and precision of the potency estimate. An endpoint of assay development is the concluded development of the assay procedure, a protocol for the performance of the bioassay. The procedure should include enough detail so that a qualified laboratory with a trained analyst can perform the procedure in a routine analysis. A strategic part of development is a look forward toward performance maintenance. Standard operating procedures for reagent, technician qualification and calibration of working reference standard.

2.6.1 Identification of Critical reagent/ equipment

It is necessary to understand the mechanism of action of drug for the intended bioassay. Based on the MOA of drug/ product, select the reliable source during development stage for like;

1. Cell line as cell substrate for In-vitro assay/ animal species for In-vivo model assay
2. Foetal bovine serum
3. Growth media/ supplementary enzymes etc.
4. Detection dye/ Kit etc.
5. Type of plates
6. Equipment etc.,

After addressing the initial required materials during development, the lot of the material should be qualified and implemented in routine analysis.

2.6.2 Assay Layout, Blocking, and Randomization

Most cell-based assays are performed using a cell culture plate (6-, 12-, 96-, or 384-well micro titer plate). Ideally, a plate is able to provide a uniform substrate for experimental treatments in all wells, including after wash steps and incubations. However, regardless of assay conditions intended to minimize the potential for bias (e.g., good analyst technique, careful calibration of pipets, controlled incubation time, and temperature), systematic gradients on the plate, independent of experimental treatments, may be observed. These gradients may occur across rows, across columns, or from the edge to the centre of the plate and are often called plate effects. Even moderate or inconsistent plate effects should be addressed during assay development, by means of plate layout strategies, blocking, randomization, and replication.

Draft for stakeholder review

Plate effects can be evaluated in a uniformity trial in which a single experimental treatment, such as an assay concentration chosen from the middle section of the concentration–response curve, is used in all wells of the plate. Another common plate effect is a differential cell-growth pattern in which the outer wells of the plate grow cells in such a way that the assay signal is attenuated. This is such a persistent problem that the choice is often made to not use the outer wells of the assay plate. Because location effects are so common, designs that place replicates (e.g., of sample by concentration combinations) in adjacent wells should be avoided.

Blocking is the grouping of related experimental units in experimental designs. Blocks may consist of individual 96-well plates, sections of 96-well plates, or 96-well plates grouped by analyst, day, or batch of cells. The goal is to isolate any systematic effects so that they do not obscure the effects 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) are applied to experimental units for that factor within a single block. 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.

Randomization is a process of assignment of treatment to experimental units based on chance so that all such experimental units have an equal chance of receiving a given treatment. Although challenging in practice, randomization of experimental treatments has been advocated as the best approach to minimizing assay bias or, more accurately, to protecting the assay results from known and unknown sources of bias by converting bias into variance.

While randomization of samples and concentrations to individual plate wells may not be practical, a plate layout can be designed to minimize plate effects by alternating sample positions across plates and the pattern of dilutions within and across plates. Where multiple plates are required in an assay, the plate layout design should, at a minimum, alternate sample positions across plates within an assay run to accommodate possible bias introduced by the analyst or equipment on a given day. It is prudent to use a balanced rotation of layouts on plates so that the collection of replicates (each of which uses a different layout) provides some protection against likely sources of bias.

A layout that provides some protection from plate effects and can be performed manually is a strip-plot design, samples are randomized to rows of a plate and dilution series are performed

Draft for stakeholder review

in different directions in different sections (blocks) on the plate to mitigate bias across columns of the plate. An added advantage of the strip-plot design is the ability to detect location effects by the interaction of sample and dilution direction (left-to-right or right-to-left). Assay variance may then be addressed, as necessary, by increased assay replication (increased number of plates in an assay).

A split-plot design, an alternative that assigns samples to plate rows randomly and randomizes dilutions (concentrations) within each row. Such a strategy may be difficult to implement even with the use of robotics.

2.6.3 Dilution Strategy

Assay concentrations of a Test sample and the Standard can be obtained in different ways. Laboratories often perform serial dilutions, in which each dilution is prepared from the previous one, in succession. Alternatively, the laboratory may prepare wholly independent dilutions from the Test sample and Standard to obtain independent concentration series. These two strategies result in the same nominal concentrations, but they have different properties related to error. Serial dilutions are subject to propagation of error across the dilution series, and a dilution error made at an early dilution will result in correlated, non-independent observations. Correlations may also be introduced by use of multichannel pipets. Independent dilutions help mitigate the bias resulting from dilution errors. It is noteworthy that when working to improve precision, the biggest reductions in variance come when replicating at the highest possible levels of nested random effects. This is particularly effective when these highest levels are sources of variability.

2.6.4 Design of Experiments

Bioassay development proceeds through a series of experiments in which conditions and levels of assay factors are varied to identify those that support a reliable and robust bioassay qualified for routine use. Those experiments may be conducted one factor at a time (OFAT), studying each parameter separately to identify ideal conditions, or through the use of multi-factor design of experiments (DOE). DOE is an efficient and effective strategy for developing a bioassay and improving bioassay performance, thus helping to obtain a measurement system

Draft for stakeholder review

that meets its requirements. In comparison to OFAT, DOE generally requires fewer experiments and also provides insight into interactions of factors that affect bioassay performance. Assay development using DOE may proceed through a series of steps: response and risk analysis; screening; response optimization; and confirmation.

2.6.5 Response and risk analysis

Bioassay optimization may begin with a systematic examination and risk assessment to identify those factors that may influence bioassay response. It is useful to visualize bioassay factors using a bioassay process map such as a cause-and-effect or fishbone diagram. Using the process map as a guide, the laboratory can examine assay factors that might affect assay performance, such as buffer pH, incubation temperature, and incubation time. Historical experience with one or several of the bioassay steps, along with sound scientific judgment, can identify key factors that require further evaluation. One tool that may be used to prioritize factors is a failure mode and effects analysis. Factors are typically scored by the combination of their potential to influence assay response and the likelihood that they will occur. The laboratory must be careful to recognize potential interactions between assay factors.

2.6.6 Screening and response optimization

Once potential key factors have been identified from response and risk analysis, the laboratory may conduct an initial screening experiment to probe for effects that may require control. Screening designs such as factorial and fractional factorial designs are commonly used for this purpose. Software is available to assist the practitioner in the selection of the design and in subsequent analysis. Analysts should take care, however, to understand their assumptions about design selection and analysis to ensure accurate identification of experimental factors.

A screening design will usually detect a few important factors from among those studied. Such factors can be further studied in a response-optimization design. Response-optimization designs such as central composite designs are performed to determine optimal settings for combinations of bioassay factors for achieving desired response. The information obtained from response optimization may be depicted as a response surface and can be used to establish ranges that yield acceptable assay performance and will be incorporated into the bioassay procedure. Establishing a true design space for a bioassay is challenging; some but

Draft for stakeholder review

not all factors and levels of random factors will be included in the development DOE, and there is no assurance that the design space is not sensitive to unstudied random factors. Similarly, there is little assurance that the assay (design space) is robust to random factors that are studied using small samples (or non-random samples of levels). Elements of DOE that may be considered include the use of blocks; deliberate confounding among interactions that are of lower interest, or known to be unimportant; robust design (response surface designs with random effects); and use of split-plot, strip-plot, or split-lot designs.

2.6.7 Confirmation

The Assay Confirmation can take the form of a qualification trial in which the assay is performed, preferably multiple independent times using optimal values for factors. Alternatively, the laboratory may determine that the bioassay has been adequately developed and may move to validation. Qualification is a good practice, not a regulatory requirement. The decision to perform confirmatory qualifying runs or to proceed to validation depends upon the strength of the accumulated information obtained throughout development.

2.7 DATA ANALYSIS DURING ASSAY DEVELOPMENT

Analysis of bioassay data during assay development enables analysts to make decisions regarding the statistical model that will be used for routine analysis, including transformation and/or weighting of data, and the development of system and sample suitability criteria.

The analysis also provides information regarding which elements of design structure should be used during outlier detection and the fitting of a full model. This may also include a plan for choosing subsets of data, such as a linear portion, for analysis or, for nonlinear bioassays, a model reduction strategy for samples similar to Standard. Once these decisions are made and proven sound during validation, they don't need to be reassessed with each performance of the assay. A process approach to enabling these decisions follows.

Step 1: Choose an appropriate statistical model. Given the complexity of bioassays and the motivation to use an approach proven reliable, fairly standardized analytical models are common in the field of bioassay analysis.

Step 2: Fit the chosen statistical model to the data without the assumption of parallelism, and then assess the distribution of the residuals, specifically examining them for departures from normality and constant variance. Transform the data as necessary or, if needed, choose a weighting scheme.

Draft for stakeholder review

Step 3: Screen for outliers, and remove as is appropriate. This step normally follows the initial choice of a suitable transformation and/or weighting method. Ideally the model used for outlier detection contains the important elements of the assay design structure, allows nonsimilar curves, and makes fewer assumptions about the functional shape of the concentration–response curve than did the model used to assess similarity. It is necessary to screen the raw data for outliers before attempting to fit the model. During assay development, a strategy should be developed for the investigation and treatment of an outlier observation, including any limits on how many outliers are acceptable. Include these instructions in the assay SOP. Good practice includes recording the process of an investigation; outlier test(s) applied, and results there from. Note that outlier procedures must be considered apart 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 (e.g., a potential OOS result). Removing data as outliers should be rare. If many values from a run are removed as outliers, that run should be considered suspect.

Step 4: Refit the model with the transformation and/or weighting previously imposed (Step 2) without the observations identified as outliers (Step 3) and re-assess the appropriateness of the model.

Step 5: If necessary or desired, choose a scheme for identifying subsets of data to use for potency estimation, whether the model is linear or nonlinear

Step 6: Calculate a relative potency estimate by analyzing the Test and Standard data together using a model constrained to have parallel lines or curves, or equal intercepts.

3.0 BIOASSAY VALIDATION

Good manufacturing practice requires that test methods used for assessing compliance of pharmaceutical products with quality requirements should meet appropriate standards for accuracy and reliability. 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. USP general chapter Validation of Compendial Procedures <1225> and ICH Q2 (R1) describe the assay performance characteristics (parameters) that should be evaluated for

Draft for stakeholder review

procedures supporting small-molecule pharmaceuticals. Although evaluation of these validation parameters is straightforward for many types of analytical procedures for well-characterized, chemically-based drug products, their interpretation and applicability for some types of bioassays has not been clearly delineated. This chapter addresses bioassay validation from the point of view of the measurement of activity rather than mass or other physicochemical measurements, with the purpose of aligning bioassay performance characteristics with uses of bioassays in practice. Assessment of bioassay performance is a continuous process, but bioassay validation should be performed when development has been completed. Bioassay validation is guided by a validation protocol describing the goals and design of the validation study. This chapter provides validation goals pertaining to relative potency bioassays. Relative potency bioassays are based on a comparison of bioassay responses for a Test sample to those of a designated Standard that provides a quantitative measure of the Test bioactivity relative to that of the Standard.

Validation parameters discussed include relative accuracy, specificity, intermediate precision, and range. Laboratories may use dilutional linearity to verify the relative accuracy and range of the method. Although robustness is not a requirement for validation, this chapter recommends that a bioassay's robustness be assessed prior to validation. In addition, describes approaches for validation design (sample selection and replication strategy), validation acceptance criteria, data analysis and interpretation, and finally bioassay performance monitoring through quality control. Documentation of bioassay validation results is also discussed, with reference to pre-validation experiments performed to optimize bioassay performance

3.1 Concept of Bioassay Validation

The bioassay validation is a protocol-driven study that demonstrates that the procedure is fit for use. A stage-wise approach to validation may be considered, as in a “suitable for intended use” validation to support release of clinical trial material, and a final, comprehensive validation prior to any regulatory filing. Preliminary system and sample suitability controls should be established and clearly described in the assay procedure; these may be finalized based on additional experience gained in the validation exercise.

The goal of bioassay validation is to confirm that the operating characteristics of the procedure are such that the procedure is suitable for its intended use.

Draft for stakeholder review

Specificity: Demonstrating specificity (also known as selectivity) requires evidence of lack of influence from matrix components such as manufacturing process components or degradation products so that measurements quantify the target molecule only. Other analytical methods may complement a bioassay in measuring or identifying other components in a sample.

Relative accuracy & Range: In order to establish the relative accuracy and range of the bioassay, validation Test samples may be constructed using a dilution series of the Standard to assess dilutional linearity (linearity of the relationship between known and measured relative potency).

Intermediate Precision (IP): The potential influences on the bioassay from inter-run factors such as multiple analysts, instruments, or reagent sources, the design of the bioassay validation should include consideration of these factors. The variability of potency from these combined elements defines the intermediate precision (IP) of the bioassay. An appropriate study of the variability of the potency values obtained, including the impact of intra-assay and inter-run factors, can help the laboratory confirm an adequate testing strategy and forecast the inherent variability of the reportable value (which may be the average of multiple potency determinations). Variability estimates can also be utilized to establish the sizes of differences (fold difference) that can be distinguished between samples tested in the bioassay.

Robustness: Robustness studies are usually performed during bioassay development, key factors in these studies such as incubation time and temperature and, for cell-based bioassays, cell passage number and cell number may be included in the validation, particularly if they interact with another factor that is introduced during the validation (e.g., a temperature sensitive reagent that varies in its sensitivity from lot-to-lot).

3.2 Bioassay Validation Protocol

A bioassay validation protocol should include the number and types of samples that will be studied in the validation; the study design, including inter-run and intra-run factors; the replication strategy; the intended validation parameters and justified target acceptance criteria for each parameter; and a proposed data-analysis plan. In addition, assay, run, and sample acceptance criteria such as system suitability and similarity should be specified before performing the validation. Depending on the extent of development of the bioassay, these may be proposed as tentative and can be updated with data from the validation.

Assay, run, or sample failures may be reassessed according to criteria which have been defined in the validation protocol and, with sound justification, included in the overall

Draft for stakeholder review

validation assessment. Additional validation trials may be required in order to support changes to the method. The bioassay validation protocol should include target acceptance criteria for the proposed validation parameters. Steps to be taken upon failure to meet a target acceptance criterion should be specified in the validation protocol, and may result in a limit on the range of potencies that can be measured in the bioassay or a modification to the replication strategy in the bioassay procedure.

3.3 Documentation of Bioassay Validation Results

Bioassay validation results should be documented in a bioassay validation report. The validation report should support the conclusion that the method is fit for use or should indicate corrective action (such as an increase in the replication strategy) that will be undertaken to generate sufficiently reliable results to achieve fitness for use. The report could include the raw data and intermediate results (e.g., variance component estimates should be provided in addition to overall intermediate precision) which would facilitate reproduction of the bioassay validation analysis by an independent reviewer. Estimates of validation parameters should be reported at each level and overall as appropriate. Deviations from the validation protocol should be documented with justification. The conclusions from the study should be clearly described with references to follow-up action as necessary. Follow-up action can include amendment of system or sample suitability criteria or modification of the bioassay replication strategy. Reference to pre-validation experiments may be included as part of the validation study report. Pre-validation experiments may include robustness experiments, where bioassay parameters have been identified and ranges have been established for significant parameters, and also may include qualification experiments, where the final procedure has been performed to confirm satisfactory performance in routine operation. Conclusions from pre-validation and qualification experiments performed during development contribute to the description of the operating characteristics of the bioassay procedure.

3.4 Bioassay Validation Design

The biological assay validation should include samples that are representative of materials that will be tested in the bioassay and should effectively establish the performance characteristics of the procedure. For relative accuracy, sample relative potency levels that bracket the range of potencies that may be tested in the bioassay should be used. Thus

Draft for stakeholder review

samples that span a wide range of potencies might be studied for a drug or biological with a wide specification range or for a product that is inherently unstable, but a narrower range can be used for a more durable product.

A minimum of three potency levels is required, and five are recommended for a reliable assessment. If the validation criteria for relative accuracy and IP are satisfied, the potency levels chosen will constitute the range of the bioassay. A limited range will result from levels that fail to meet their target acceptance criteria. Samples may also be generated for the bioassay validation by stressing a sample to a level that might be observed in routine practice (i.e., stability investigations). Additionally, the influences of the sample matrix (excipients, process constituents, or combination components) can be studied strategically by intentionally varying these together with the target analyte, using a multifactorial approach. Often this will have been done during development, prior to generating release and stability data. The bioassay validation design should consider all facets of the measurement process. Sources of bioassay measurement variability include sample preparation, intra-run factors, and inter-run factors. Representative estimation of bioassay variability necessitates consideration of these factors. Test sample and Standard preparation should be performed independently during each validation run. The replication strategy used in the validation should reflect knowledge of the factors that might influence the measurement of potency. Intra-run variability may be affected by bioassay operating factors that are usually set during development (temperature, pH, incubation times, etc.); by the bioassay design (number of animals, number of dilutions, replicates per dilution, dilution spacing, etc.); by the assay acceptance and sample acceptance criteria; and by the statistical analysis (where the primary endpoints are the similarity assessment for each sample and potency estimates for the reference samples). Operating restrictions and bioassay design (intra- and inter-run formulae that result in a reportable value for a test material) are usually specified during development and may become a part of the bioassay operating procedure. IP is studied by independent runs of the procedure, perhaps using an experimental design that alters those factors that may have an impact on the performance of the procedure. Experiments (including those that implement formalized design of experiments [DOE]) with nested or crossed design structure can reveal important sources of variability in the procedure, as well as ensure a representative estimate of long-term variability.

During the validation it is not necessary to employ the format required to achieve the reportable value for a Test sample. A well-designed validation experiment that combines both

Draft for stakeholder review

intra-run and inter-run sources of variability provides estimates of independent components of the bioassay variability. These components can be used to verify or forecast the variability of the bioassay format. A thorough analysis of the validation data should include graphical and statistical summaries of the validation parameters' results and their conformance to target acceptance criteria. The analysis should follow the specifics of the data-analysis plan outlined in the validation protocol. In most cases, log relative potency should be analyzed in order to satisfy the assumptions of the statistical methods.

3.5 Validation Strategies for Bioassay Performance Characteristics

Parameters that should be verified in a bioassay are relative accuracy, specificity, IP (which incorporates repeatability), and range. Other parameters discussed in general chapter <1225> and ICH Q2 (R1) such as detection limit and quantitation limit has not been included because they are usually not relevant to a bioassay that reports relative potency. Linearity is not part of bioassay validation, except as it relates to relative accuracy (dilutional linearity). There follow strategies for addressing bioassay validation parameters.

3.5.1 Specificity

For products matrices and different but related biopharmaceutical molecules, specificity involves demonstrating no significant interference from matrix components or different but related biopharmaceutical molecules. This can be assessed via parallel dilution of the Standard and matrix or different but related biopharmaceutical molecules, if there is no significant response for matrix or different but related biopharmaceutical molecules and the response confirms to the no interference from matrix or different but related biopharmaceutical molecules.

For identical assay, response can be assessed via parallel dilution of the Standard and identical sample if the curves are similar and the potency conforms to expectations of a Standard-to-Sample comparison, the bioassay is specific against the compound. For these assessments both similarity and potency may be assessed using appropriate equivalence tests.

3.5.2 Relative Accuracy (Dilutional Linearity)

The relative accuracy of a relative potency bioassay is the relationship between measured relative potency and known relative potency. Relative accuracy in bioassay refers to a unit slope (slope = 1) between log measured relative potency and log known relative potency. The

Draft for stakeholder review

most common approach to demonstrating relative accuracy for relative potency bioassays is by construction of target potencies by dilution of the standard material or a Test sample with known potency. This type of study is often referred to as a dilutional linearity study. The results from a dilutional linearity study should be assessed using the estimated relative bias at individual levels and via a trend in relative bias across levels. The relative bias at individual levels is calculated as follows:

$$\text{Relative bias} = \left(\frac{\text{Measured Potency}}{\text{Target Poency}} - 1 \right) \%$$

The trend in bias is measured by the estimated slope of log measured potency versus log target potency, which should be held to a target acceptance criterion. If there is no trend in relative bias across levels, the estimated relative bias at each level can be held to a prespecified target acceptance criterion that has been defined in the validation protocol.

3.5.3 Intermediate Precision

Because of potential influences on the bioassay by factors such as analysts, instruments, or reagent lots, the design of the bioassay validation should include evaluation of these factors. The overall variability from measurements taken under a variety of normal test conditions within one laboratory defines the IP of the bioassay. IP is the ICH and USP term for what is also commonly referred to as inter-run variability. IP measures the influence of factors that will vary over time after the bioassay is implemented. These influences are generally unavoidable and include factors like change in personnel (new analysts), receipt of new reagent lots, etc.

When the validation has been planned using multifactor DOE, the impact of each factor can first be explored graphically to establish important contributions to potency variability. The identification of important factors should lead to procedures that seek to control their effects, such as further restrictions on intra-assay operating conditions or strategic qualification procedures on inter-run factors such as analysts, instruments, and reagent lots. Contributions of validation study factors to the overall IP of the bioassay can be determined by performing a variance component analysis on the validation results. Variance component analysis is best carried out using a statistical software package that is capable of performing a mixed-model analysis with restricted maximum likelihood estimation (REML). A variance component analysis yields variance component estimates such as corresponding to intra-run and inter-run variation.

Draft for stakeholder review

$$\%CV \text{ (intra assay)} = \frac{SD}{Mean} \times 100$$

And

$$\%CV \text{ (inter assay)} = \frac{SD}{Mean} \times 100$$

These can be used to estimate the IP of the bioassay, as well as the variability of the reportable value for different bioassay formats (format variability). IP expressed as percent coefficient of variation (%CV).

This formula can be used to determine a testing format suitable for various uses of the bioassay (e.g., release testing and stability evaluation).

3.5.4 Range

The range of the bioassay is defined as the true or known potencies for which it has been demonstrated that the analytical procedure has a suitable level of relative accuracy and IP. The range is normally derived from the dilutional linearity study and minimally should cover the product specification range for potency. For stability testing and to minimize having to dilute or concentrate hyper- or hypo-potent Test samples into the bioassay range, there is value in validating the bioassay over a broader range.

4.0 BIOASSAY MAINTENANCE AND ASSAY TRANSFER

Once a bioassay has been validated it can be implemented. However, it is important to monitor its behavior over time. This is most easily accomplished by maintaining statistical process control (SPC) charts for suitable parameters of the Standard response curve and potency of assay QC samples. The purpose of these charts is to identify at an early stage any shift or drift in the bioassay. If a trend is observed in any SPC chart, the reason for the trend should be identified. If the resolution requires a modification to the bioassay or if a serious modification of the bioassay has occurred for other reasons (for example, a major technology change), the modified bioassay should be revalidated or linked to the original bioassay by an adequately designed bridging study with acceptance criteria that use equivalence testing.

The development and validation of a bioassay, though discrete operations, lead to ongoing activities. Assay improvements may be implemented as technologies change, as the laboratory becomes more skilled with the procedure, and as changes to bioassay methodology require re-evaluation of bioassay performance. Some of these changes may be responses to unexpected performance during routine processing. Corrective action should be monitored

Draft for stakeholder review

using routine control procedures. Substantial changes may require a study verifying that the bioassay remains fit for use. An equivalence testing approach can be used to show that the change has resulted in acceptable performance. A statistically-oriented study can be performed to demonstrate that the change does not compromise the previously acceptable performance characteristics of the assay.

Assay transfer assumes both a known intended use of the bioassay in the recipient lab and the associated required capability for the assay system. These implicitly, though perhaps not precisely, demarcate the limits on the amount of bias and loss of precision allowed between labs. Using two laboratories interchangeably to support one product will require considering the variation between labs in addition to intermediate precision for sample size requirements to determine process capability.

5.0 ANALYSIS OF BIOASSAY

Most of the potency determination assays are relied on biological assay where the chemical characterization assay does not mimic the exact mechanism of action of the drug for many products, bioassays are still essential for the determination of potency and the assurance of activity of many proteins, vaccines, complex mixtures, and products for cell and gene therapy, as well as for their role in monitoring the stability of biological products. In biological assays, the variable is the amount of drug, the system is a cell, and a response is a property of the cell that changes when the drug is given. The experimental design employed in biological assays is different from those typical used in chemical and physical experiments since the intrinsic variation in living matter increases the complexity of the experiment and the data analysis. The experimental design has not only to take into account the intrinsic variability in the biological test units but also other factors as operators, day variation, batch variation, etc.

5.1 Principle of Biological assay

A biological assay or bioassay is used to determine the effect of a substance/product on a certain type of living matter. A biological test system, for example animals, tissue, etc, is exposed to a particular stimulus like a drug, whose concentration (dose) is usually varied. The magnitude of the response of the biological system depends on the dose. In contrast with physical or chemical methods, detailed information of the drug activity as a function of the

Draft for stakeholder review

dose is obtained. A special characteristic of bioassays is that one of the largest sources of variation in the outcome is the difference between the test units, and since the response is dependent on living matter this introduces large variability between measurements obtained by identical operations. The response can be a characteristic like a count, number, body weight of the test units, or a change in a characteristic, change in body weight, or the occurrence of a certain phenomena, e.g. viability or death.

Biological assays are usually comparative. The capacity of a substance to cause a specific effect is estimated relative to a standard. The standard and the test preparations are identical in their biological activity principle and differ only in extends to which they are diluted by solvents. This type of bioassay is used in the pharmaceutical industry to determine the potency of test product relative to a standard sample.

5.2 Introductory step for analysis of Biological assay

Ensure the method development & validation activity should be completed and the preliminary information of the assay is fixed to avoid further decision and changes and also it will help in investigation of any nonconformity.

Analysis of biological assay should be analyse base on protocol where

- 5.2.1 Cell line handling and passage range are the most critical step for biological assay. Select the validated range of passage for analysis, where the responses are reproducible for shape, EC_{50} , range of upper asymptote and lower asymptote.
- 5.2.2 Based on assay model, for linear model select a linear region (Choosing model) of the curve for the standard and sample (subset of data) to be used in the determination of the relative potency using the predefined scheme. Exclude only data known to result from technical problems such as contaminated wells, non-monotonic concentration–response curves, etc.
- 5.2.3 Screen the data for detection of potential outliers, as chosen during development or as per defined procedure, including any weighting and transformation. This is done first without assuming similarity of the Test and Standard curves but should include important elements of the design structure, ideally using a model that

Draft for stakeholder review

makes fewer assumptions about the functional form of the response than the model used to assess similarity.

- 5.2.4 Determine which potential outliers are to be removed and fit the model to be used for System suitability assessment. The approach for the outlier removal should be fixed during development and shall be part of the procedure for maximum valid outlier in assay curve etc. If a large number of outliers are found above that expected from the defined procedure, that calls into question the assay. After removal of outlier, the assay does meet the system suitability; invalid the assay and repeat the assay based on justification for the occurrence.
- 5.2.5 System suitability: System suitability assesses whether the assay Standard preparation and any controls behaved in a manner consistent with past performance of the assay. If an assay (or a run) fails system suitability, the entire assay (or run) is discarded and no results are reported other than that the assay (or run) failed. Assessment of system suitability usually includes adequacy of the fit of the model used to assess similarity. For linear models, adequacy of the model may include assessment of the linearity of the Standard curve. If the suitability criterion for linearity of the Standard is not met, the exclusion of one or more extreme concentrations may result in the criterion being met. The significant level for linearity must be evaluated as most precise result may fail at lower significant level ($P=0.05$) i.e. 95%, increase the significant level to ($P=0.0001$) 99.99% may serve the problem. Examples of other possible system suitability criteria include background, positive controls, max/min, max/background, slope, IC_{50} (or EC_{50}), and variation around the fitted model.
- 5.2.6 Sample suitability for each Test sample. This is done to confirm that the data for each Test sample satisfy necessary assumptions. If a Test sample fails sample suitability, results for that sample are reported as “Fails Sample Suitability.” Relative potencies for other test samples in the assay may still be reported. Most prominent of sample suitability criteria is similarity, whether parallelism for parallel models or equivalence of intercepts for slope-ratio models. For nonlinear

Draft for stakeholder review

models, similarity assessment involves all curve parameters other than EC_{50} (or IC_{50}).

- 5.2.7 Once the Assay system suitability and sample suitability meet the (i.e., sufficiently similar concentration–response curves or similar straight-line subsets of concentrations), calculate relative potency estimates assuming similarity between Test and Standard, i.e., by analyzing the Test and Standard data together using a model constrained to have exactly parallel lines or curves, or equal intercepts.
- 5.2.8 If the assay are more variable and often a single assay is not sufficient to achieve a reportable value, and potency results from multiple assays can be combined into a single potency estimate. Repeat steps 1–6 multiple times, as specified in the assay procedure or monograph, before determining a final estimate of potency and a confidence interval.

For replacement of missing data, most modern statistical methodology and software do not require equal numbers at each combination of concentration and sample. Thus, unless otherwise directed by a specific monograph or method analysts generally do not need to replace missing values.

5.3 Statistical Analysis Models

Many statistical tools computer based/ statistical software can be successfully used to describe a concentration–dose response relationship. The first consideration in choosing a model is the form of the assay response. Is it a number, a count, or a category such as Dead/Alive? The form will identify the possible models that can be considered. Other considerations in choosing a model include the need to incorporate design elements in the model and the possible benefits of means models compared to regression models.

This section presumes that decisions were made to fix statistical model during development, checked during validation, and then use for routine analysis.

6.0 IMPROVING OR UPDATING A BIOASSAY SYSTEM

Draft for stakeholder review

A new version of a bioassay may improve the quality of bias, precision, range, robustness, specificity, lower the operating costs or offer other compelling advantages. When improving or updating a bioassay system a bridging study may be used to compare the performance of the new to the established assay. A wide variety of samples (e.g., lot release, stability, stressed, critical isoforms) can be used for demonstrating equivalence of estimated potencies. Even though the assay systems may be quite different (e.g., an animal bioassay versus a cell-based bioassay), if the assays use the same Standard and mechanism of action, comparable potencies may reasonably be expected. If the new assay uses a different Standard, the minimum requirement for an acceptable comparison is a unit slope of the log linear relationship between the estimated potencies. An important implication of this recommendation is that poor precision or biased assays used early can have lasting impact on the replication requirements, even if the assay is later replaced by an improved.