Bioassays 〈*G3-18-190*〉

(バイオアッセイ 〈*G3-18-190*〉)

1. Role of Biological Activity Evaluation to Quality Assurance of Biotechnological Products (Biopharmaceuticals)

Evaluating the biological activity through biological assays is required to confirm that biopharmaceuticals have intended properties, in addition to evaluating the structure and physicochemical properties. This is because that an active ingredient of biopharmaceutical is a large molecule having complex structure, and a mixture of various molecular species, making it difficult to determine the higher-order structure by physicochemical analysis. To determine biological activity, it is necessary to conduct biological activity tests (bioassays) using biological test methods, which need to be designed by considering the mechanism of action of the drug being evaluated.

In bioassays, an appropriate analysis model is applied to the dose-response relationship of both the reference material and the test sample, and the relative activity (relative potency) is calculated by comparing the parameters of the obtained regression equation. If the reference material and the test samples contain the same active ingredient, it is believed that the biological response showing the same dose-response relationship will be obtained, so the calculation of relative potency assumes that the dose-response relationships of the reference material and test samples are in agreement. The dose-response agreement is generally determined by statistically evaluating the similarity of the regression line or regression curve. Similarity is sometimes expressed as parallelism in the evaluation of parallel line models and nonlinearity models.

This general information provides important notices to properly perform bioassays, including analysis methods such as the calculation of relative potency, standard approaches to the analytical procedure validation and the establishment of suitability testing conditions, as well as considerations in the development and life cycle management of bioassays.

2. Calculation of Relative Potency

The quantitative measure of biological activity based on characteristics related to the biological properties of a drug is usually "potency", which can be expressed as a "unit". In principle, the unit is defined as a relative value to a pre-determined reference standard. However, the unit may also be defined based on absolute standards, such as the substrate degradation capacity per unit time for enzyme activity. The obtained value can be expressed as potency per volume (units/mL) or potency per amount of protein (units/mg). Potency per amount of protein is called specific activity. In the case of a drug for which the unit is not assigned, the relative potency (%) can be calculated as the relative activity to the reference standard or the reference material. Hereinafter, the term "reference material" will be used, including the case where reference standard is used.

In bioassays to determine relative potency by comparison with the reference material, an appropriate analysis model is usually applied to the dose-response relationship for the reference material and test sample, and the relative potency is calculated by comparing the parameters of the obtained regression equations. The main analysis models used are 1) Nonlinear model, 2) parallel-line model, and 3) Slope-ratio concentration-response model.

Fig. 1 indicates a typical procedure for analyzing assay data. In the analysis of bioassay data, after selecting data to be used for analysis by statistical processing, etc., an analytical model is applied individually to the regression curves of the reference material and test sample (unconstrained model), and the validity of the test is determined based on the indicators for model suitability and an evaluation of similarity (see 4. Judgment of Suitability testing for details). If the result is valid, an analysis model where the parameters of the regression curves of the reference material and test sample are the same (constrained model) is applied, and the relative potency is calculated using the obtained regression equation. To reduce the effects of variation, the relative potencies obtained from multiple analytical runs are integrated as necessary and calculated as the reported value. An analytical run of a bioassay is considered to be one set of samples consisting of a serial dilution series of the test sample and the reference material, including appropriate replications to calculate the relative potency. In some cases, an analytical run requires only one microplate (plate), while in other cases multiple plates are required.

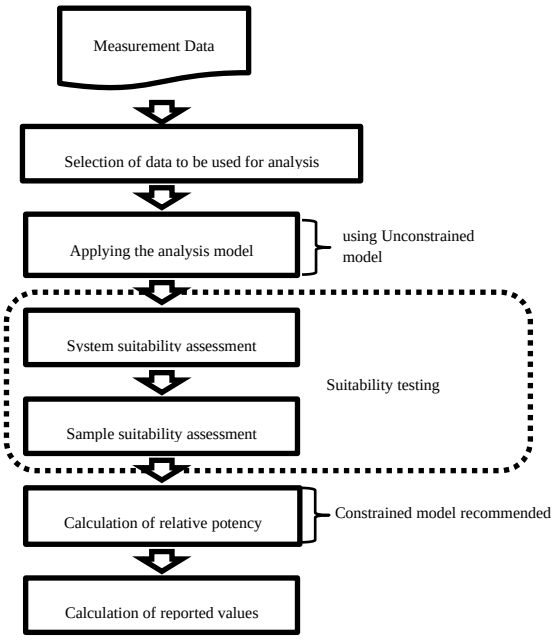

Fig.1 Analysis Procedures using Unconstrained model and Constrained model

The calculation of relative potency is premised on the similarity of dose-response between the reference material and test sample (see 4. Suitability testing for details). For example, in the case of a four-parameter logistic model, assuming that parallelism is maintained, i.e., that the upper and lower asymptotes and slope parameters of the dose curve of the reference material and those of the test sample are the same, the relative potency can be calculated by comparing the EC_{50} values of both. A model that uses a regression equation with matched upper and lower asymptotes and slope parameters of the regression curves between the reference material and test sample is called a constrained model (reduced model, restricted model). On the other hand, there is a method that uses individual regression equations without matching the upper and lower asymptotes and slope parameters of the regression curves between the reference material and test sample and is called an unconstrained model (full model). As mentioned above, since the calculation of relative potency is premised on parallelism, it is generally recommended to use a constrained model that calculates the relative potency using a regression curve with common asymptotes. If the shapes of the dose-response curves of the reference material and test sample are sufficiently similar, it may be possible to use an unconstrained model. It is necessary to decide in advance which method to use. There may be a large discrepancy between the analysis results (EC_{50} ratio) of the constrained model and the unconstrained model for some data, but such test results can be invalidated by appropriately defining the test validity conditions.

2.1. Nonlinear model

Nonlinear dose-response models are generally S-shaped functions, and are used when the concentration range is wide enough that the response is limited by upper and lower asymptotes. The most widely used model among these models is the four-parameter logistic model.

When the observed response is y, the concentration is z, and the error term is e, the four-parameter logistic model is expressed as follows:

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

Upper asymptote : A
Lower asymptote : D
Slope parameter : B
50% effective concentration (EC_{50}) : C

B is not a coefficient that indicates the slope of curve at EC_{50}, but the slope of curve at EC_{50} is expressed as S below. It should be noted that S is the slope with respect to concentration, and its value may be different from the apparent slope of the sigmoid curve where concentrations are transformed logarithmically.

$$S = -\frac{1}{4} \times \frac{A-D}{C} \times B$$

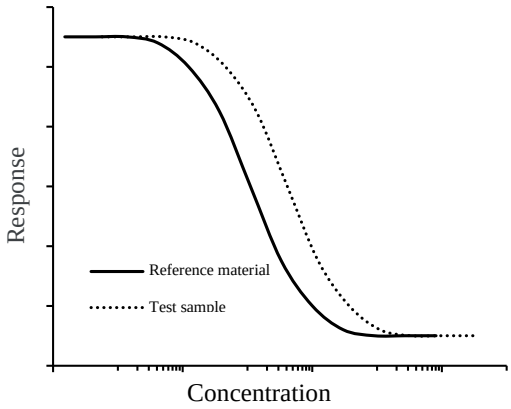

Fig.2 Example of parallel curves obtained using a nonlinear curve model

When the shapes of the two graphs are consistent, including the agreement of the upper and lower asymptotes, it is possible to calculate the relative potency, which is the ratio of EC_{50} (reference material/test sample), represented by the parameter C.

2.2. Parallel-line model

If the dose-response relationship is a linear function between the log z of the concentration x and the response y, a parallel-line model can be used. If the general concentration-response model is transformed to linear one with the equation $x = \log(z)$, it can be expressed as follows, where y-intercept is α and slope is β:

$$y = \alpha + \beta\log(z) + \mathrm{e} = \alpha + \beta x + \mathrm{e}$$

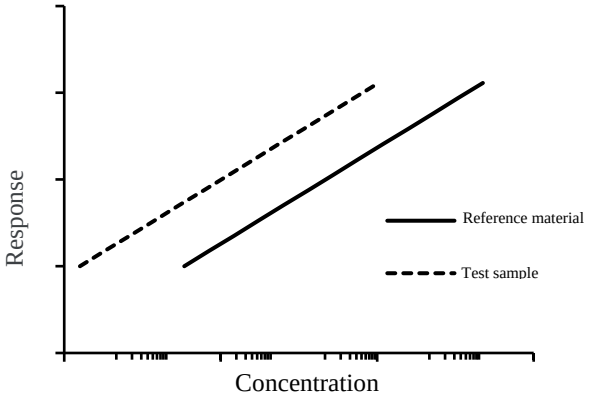

Fig.3 Example of the parallel-line model

When the dose-response lines are parallel, the horizontal shift indicates the difference in the biological activity of

the object being measured. This horizontal difference is the logarithm of the relative potency, log(ρ), which is calculated as the difference in the intercepts of lines of the test sample and reference material divided by the slope β. Thus, the relative potency is expressed by the following equation (T is the test sample and S is the reference material). The reported value may be calculated by combining the relative potencies obtained from multiple analytical runs.

$$\rho = \text{antilog}\left(\frac{\alpha_T - \alpha_S}{\beta}\right)$$

2.3. Slope-ratio concentration-response model

If the dose-response data is well fitted by linear regression, a slope-ratio concentration-response model can be used.

The slope-ratio concentration-response model assuming similarity can be expressed as follows:

$$Y_S = \alpha + \beta z + e = \alpha + \beta_S z + e$$

$$Y_T = \alpha + \beta(\rho z) + e = \alpha + \beta_S \rho z + e = \alpha + \beta_T z + e$$

The characteristic of slope-ratio model is that the intercepts of lines with different potencies are the same, but their slopes are different. The common intercept need not be at the origin.

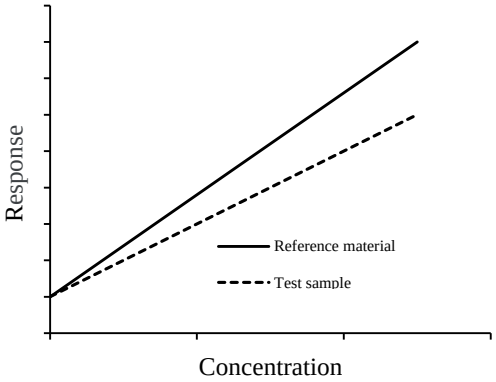

Fig.4 Example of the slope-ratio concentration-response model

A test method using the slope-ratio concentration-response model includes, at a minimum, measurements of the reference material and test sample at one or more concentration points each, as well as measurement of a blank sample. Since concentrations are not log-transformed, concentrations are generally plotted at equal intervals on a linear scale rather than a logarithmic scale. The model consists of a common intercept, a slope based on the test sample results, and a slope based on the reference material results, and the relative potency is calculated from the ratio of slopes as follows. The reported value may be calculated by combining the relative potencies obtained from multiple analytical runs.

Relative potency
=slope of the test sample／slope of the reference material
$=\beta\rho / \beta = \rho$

3. Validation of Analytical Procedures

Performance characteristics that need to be considered during analytical procedure validation of a bioassay are typically accuracy, linearity, specificity, intermediate precision (including repeatability), and range. Robustness is not a requirement for analytical procedure validation and is evaluated during analytical procedure development. However, important factors in cell-based bioassays, such as incubation time and temperature, cell density and cell passage number used, serum components, and detection reagent lots, should be included in the validation protocol, especially when these factors are related to other factors.

3.1. Accuracy

Accuracy of a bioassay refers to the closeness of agreement between the measured relative potency (the actual measurement value) and the known relative potency (the theoretical expected value). The most common method for evaluating the accuracy of a bioassay is to calculate the relative bias from the potency obtained by measuring the diluted preparation of the reference material or the test sample with known potency and the expected potency theoretically calculated from the dilution factor. The relative bias at each concentration is calculated as the ratio of the logarithmic measured potency to the logarithmic expected potency using the following equation:

Relative bias
$=100 \times$ (measured potency／expected potency−1) %

Verify that the relative bias meets the predefined expected criterion. If there is no consistent trend in the relative bias across concentrations, the estimated relative bias at each concentration can be used as the predefined expected criterion.

3.2. Linearity

Linearity of a bioassay refers to a linear relationship between the expected potency and the measured potency within a certain concentration range. The degree of linearity can be evaluated by plotting the measured potency against the expected potency and using statistical techniques. For example, the slope of the regression line, the correlation coefficient (r), and the residual sum of squares obtained by calculating the regression line using the least squares method, are evaluated. Analyzing a graph plotting the residuals against the expected potency is also useful in evaluating linearity.

3.3. Specificity

Specificity of a bioassay refers to the ability to distinguish the active ingredient contained in the drug substance or drug product being measured from a molecule related to but different from the active ingredient, as a response from the

bioassay. In some cases, it is necessary to demonstrate the absence of interference from components related to the expected substance, such as expected product-related substances contained in forced degradation samples, as well as from matrix components that are predicted to coexist.

3.4. Intermediate Precision

Intermediate precision refers to the precision when measurements are made in the same laboratory with different conditions include days, analysts, equipment, etc. Intermediate precision is useful for evaluating the effects of factors that vary over time after the introduction of the bioassay. Since these effects are generally unavoidable, it is recommended for the validation protocol of the bioassay to include an evaluation of factors such as variations in analysts, different equipment or devices, and different reagent lots, taking into account their effects on the procedures. Variations under the same conditions can also be evaluated together as repeatability.

3.5. Range

Range of a bioassay is defined as the range of relative potencies over which the analytical procedure has been shown to have adequate levels of accuracy, intermediate precision and linearity. The range should normally include, as a minimum, the product specification range for potency. It is useful to validate a bioassay over a wider range for stability studies or to minimize the need to dilute or concentrate high or low potency samples into the bioassay range.

4. Suitability testing

In tests for calculating relative potency, it is important that the reference material and the test sample show the same dose-response relationship, and the shapes of the dose-response curves of both are confirmed to be similar by satisfying the predetermined suitability testing conditions. Relative potency can be calculated only when the curves obtained from the reference material and the test sample show similarity within the predetermined concentration range of the test method and satisfy the goodness of fit to the model required for analysis.

When establishing the specification and test methods, in addition to studies for analytical method development and verifying analytical performance through analytical procedure validation, it is necessary to establish the suitability testing conditions that correspond to system suitability in physicochemical analysis in order to ensure that daily tests are being performed appropriately.

Suitability testing conditions include 1) evaluation of system suitability and 2) evaluation of sample suitability (similarity of dose-response between the reference material and test sample). As the suitability testing conditions, it is important to select appropriate indicators to ensure the expected analytical performance and establish determination criteria. Suitability testing conditions should be considered during the development of the bioassay and established before validation.

4.1. System suitability

The test method is evaluated to ensure that it remains in an appropriate condition to perform the intended test. Dose-response curves, usually obtained with the reference materials or QC samples (samples that have been adequately characterized, including biological activity), are confirmed to produce the intended response, based on the biological activity of the preparation, by checking that several parameters are within a predefined range. For example, in the four-parameter logistic model, the coefficient of determination, the value of the upper or lower asymptote, the difference or ratio between the upper and lower asymptotes, the ratio between the maximum and minimum observed responses, the slope parameter, EC_{50}, etc. are used as indicators. It can be confirmed that a certain level of performance is maintained by determining the suitability of multiple indicators. In addition, the relative potency of the QC sample can also be established as an indicator (in this case, it is also necessary to evaluate the similarity of the dose-response curves of the reference material and the QC sample). In the parallel-line model, the difference or ratio between the maximum response and the minimum response observed, and in the slope-ratio concentration-response model, the intercept and the response of the high concentration reference material can be used as indicators that are confirmed to be within a predefined range.

4.2. Sample suitability

Sample suitability is evaluated to ensure that the dose-responses of the reference material and test sample are similar. The measurement of relative potency requires that the shapes of the dose-response curves of the reference material and test sample are in agreement, thus the sample suitability is evaluated based on the similarity of the dose-response curves of the reference material and test sample. Items to be evaluated for similarity include parameters other than EC_{50}, i.e., the upper and lower asymptotes and the slope parameter, in the four-parameter logistic model; the slope of the regression line in the parallel-line model; and the intercept in the slope-ratio concentration-response model.

Methods for evaluating similarity typically include the difference test, which tests the null hypothesis that there is no difference in the dose-response between the reference material and the test sample, or the equivalence test, which evaluates similarity by defining the acceptable range within which the dose-responses of the reference material and the test sample are deemed equivalent.

4.2.1. Difference test

Difference test uses analysis of variance to evaluate the null hypothesis that there is no difference in the dose-responses between the reference material and the test sample. The advantage of the difference test is that it is possible to

make a judgment on each individual test result and it is not necessary to accumulate data in advance to determine the acceptable range. However, it is important to note that the greater the variance in the data, the larger the residual variance, making it easier to determine that there is no difference in the dose-response relationship, i.e., that the dose-response relationships are consistent, and the smaller the variance, the easier it is to determine that the dose-responses are not consistent.

One of difference tests is a method in which the variances in the measurement results of the reference material and test sample are divided into several components, and the agreement between the two curves is evaluated based on the judgment results for the null hypotheses established for each component. In a typical method, p-values for regression (assessing whether regression works), non-parallelism (assessing the divergence between an unconstrained model and a constrained model), and non-linearity (or lack of fit, assessing the data divergence from an unconstrained model) are calculated using the analysis of variance, etc., to judge the validity of suitability testing. Another method of difference tests (extra sum of squares method) involves comparing the mean squares when the each of dose-responses for the reference material and test sample is regressed by fitting a model, and those when the both dose-responses are regressed by fitting a model using some of the parameters in common assuming that the shapes of dose-response curves of the reference material and the test sample are the same, and evaluating parallelism by using the p-values to determine whether there is a significant difference of fit between the models.

4.2.2. Equivalence test

In the equivalence test, define the indicators and their acceptance criteria for determining the similarity of the dose-response curves of the reference material and test sample, and determine that the two are equivalent, i.e., similar, when the indicators obtained in each test are within the acceptance criteria. For example, in a four-parameter logistic model, parameters other than EC_{50}, i.e., the upper and lower asymptotes, and slope parameter, can be used to test the equivalence of the dose-response curves of the reference material and test sample. The ratio of the upper asymptotes of the test sample and reference material, the ratio of the lower asymptotes of the test sample and reference material, and the ratio of the slope parameters of the test sample and reference material, etc., are used for the determination. During the development of the test methods, it is necessary to determine which parameters are important and how to evaluate the equivalence.

5. Considerations in Bioassay Development and Lifecycle Management

In developing bioassays, it is important to select an appropriate test methods and critical reagents, such as reagents and cell lines, to be used in the tests. Appropriate methods for data analysis should also be selected. It is recommended to identify operational parameters that affect test performance and understand the relationship between test performance and various operational parameters, and then set test conditions that provide the desired analytical performance. It is useful to evaluate robustness of the test to ensure that its analytical performance is consistently maintained.

5.1. Selection of Test Methods

Bioassays used in specifications and test methods are established as part of a quality control strategy. Generally, an assay system that reflects the mechanism of action of the active ingredient is used, but for products with multiple mechanisms of action, the necessary one is selected among them. In cell-based bioassays, cell proliferation and cytokine secretion, etc., which reflect the mechanism of action of the active ingredient, are used as indicators of cellular responses, and reporter assays may also be performed using cells transfected with reporter genes that respond to specific intracellular signaling. In either case, it is necessary that the relationship between the clinically expected effect and the bioactivity measured in the selected bioassay is explained properly.

5.2. Establishment of Test Conditions

Bioassay development is done through a series of experiments in which the conditions and levels of test factors are varied to identify the conditions and levels of the test factors to establish reliable and robust bioassay test conditions. These experiments are performed either by examining each factor individually to identify ideal conditions or using multifactorial experimental design. The main factors that can affect the results of cell-based bioassays include incubation temperature, incubation time, cell density, medium composition (including serum components), cell passage number, and passaging schedule. In bioassays using plates, care must be taken because the layout of samples on the plate may affect test results. Even if all wells are treated with uniform experimental treatments (washing, incubation, addition of substrate, etc.), systematic gradients unrelated to the experimental treatments may be observed on the plate. When these biases occur in all rows, columns, and from the edge to the center of the plate, they are called plate effects. During test method development, measures should be taken to minimize plate effects through means such as layout of the plate, blocking, randomization, and replication. During development, it is advisable to identify the critical factors in the test method, including sample layout on the plate, and to determine optimal values or ranges for those factors that result in acceptable performance (accuracy, precision, etc.).

5.3. Data Analysis on Test Method Development

It is useful to use the bioassay data accumulated during development to determine statistical analysis methods and suitability testing conditions. Specifically, it is necessary to determine the analytical model to be applied to the data

(nonlinear model, parallel-line model, and slope-ratio concentration-response model, etc.) and, if necessary, the method for data transformation and weighting. Regarding model appropriateness, it may be useful to fit the analytical model to the data without making the assumption of parallelism and evaluate the distribution of the residuals for deviations from normality and homoscedasticity over the entire concentration range in the test. It is also recommended that, if necessary, during the development of the test method, methods for investigating and dealing with outliers are developed and linked to the test procedure.

5.4. Control of Reference Materials

In bioassays, the relative activity of the test sample to the reference material is evaluated. In order to ensure the reliability of the reported values, it is necessary to ensure the reliability of the reference material. Therefore, reference materials should be treated as critical reagents. A reference material must represent the samples to be evaluated and be prepared from a lot whose efficacy and safety have been confirmed in nonclinical or clinical trials, or from a lot that has undergone characterization. Reference materials must be stored under conditions that maintain their potency over the period in which they are used. Reference materials may be stored under conditions (storage temperature, storage container, formulation, etc.) that differ from those of the drug substances and the drug products. For reference materials for biopharmaceuticals, periodic stability evaluations are recommended because activities may decrease due to chemical changes such as glycation and oxidation during storage, resulting in loss of similarity and decrease in potency over time. In addition, because there is a possibility of discontinuity in potency may occur when lots are updated, the method of evaluating the updated lot must be properly defined and implemented in the procedure for updating the reference material lots. In addition to the measures mentioned above, to deal with the decrease in potency of reference materials over time and discontinuity when lots are updated, it is useful to retain a portion of each reference material lot in order to investigate future equivalence with reference materials and fluctuations in potency test results. It is also useful to set up a two-stage system of the reference material, in which an initially established reference material is used as a primary reference material, and a reference material calibrated against the primary reference material is used as a working reference material for commercial production tests, and its biological activity is evaluated based on the primary reference material when the working reference material is updated.

5.5. Establishment and Control of Critical Reagents

Reagents that have a significant effect on test performance should be controlled as critical reagents. For example, in cell-based assay of antibody drugs with neutralizing activity, physiologically active substances that are neutralized by the antibody are included. For critical reagents, evaluation items for ensuring quality, their acceptance criteria, storage conditions, and lot update procedures must be established in advance, taking into consideration the quality characteristics and stability. When updating the lot, the reagent should actually be used in the bioassay to confirm that the intended results are obtained.

In cell-based bioassays, cells are included as critical reagents because cell characteristics affect test performance. To ensure the reliability of the test results, it is important to perform characterization to confirm that the cells have characteristics suitable for the test, and to properly control the cell line. To ensure a sufficient and stable supply of the cells, it is recommended to establish a cell bank system. The cell bank system should be established taking into consideration the growth characteristics of the cells, the cell passage number, the cell density used in the bioassay, and the frequency of the test. It is desirable to keep detailed records of information on the functional and genetic characteristics of the cell line used in the bioassay, including the history from the origin of the cell line to banking. For cells prepared from the cell bank and used in the test, conditions to ensure the test performance should be established, such as using the cells in the test within a predetermined range of passage numbers.

6. Glossary

Bioassays (Biological assay)： Analytical methods used to measure biological activity. Depending on the characteristics of the active ingredient, methods with indicators such as enzyme reactions, binding to target molecules, cellular responsiveness, and reactions in individual animals are used.

Relative Potency: Relative activity calculated by applying an appropriate analytical model to the dose-response relationship of the reference material and test sample, and comparing the parameters of the obtained regression equation.

Suitability testing: Suitability testing is a set of predefined criteria for determining whether a test performed is valid, and is usually composed of system suitability that defines the reactivity of the reference material, and the sample suitability that defines criteria for similarity between reactivity of the reference material and test sample. Suitability testing is equivalent to system suitability in physicochemical test, but the unique feature of suitability testing is that sample data is used to make the judgment.

Similarity: Similarity refers to the degree of agreement between the dose-response lines (curves) of the reference material and test sample. The calculation of relative potency is premised on the fact that the shapes of the dose-response lines (curves) of the reference material and test sample are similar. Similarity is sometimes expressed as parallelism in the evaluation of parallel-line models or nonlinear models.

Difference test: A method of assessing suitability testing by establishing the null hypothesis that there is no difference. If

the null hypothesis is not rejected, meaning that it cannot be said that there is a difference, it can be determined that the indicator meets the requirements for suitability testing. Typically, the significance of regression, non-parallelism (divergence between models), and non-linearity (divergence between points on the regression curve and the measured values) is tested. When using a four-parameter logistic model regression, extra sum of squares method may also be used.

Extra sum of squares method: A method for testing parallelism using analysis of variance that compares the residual variance causing from differences in models (constrained model and unconstrained model) with the residual variance when regressing using the unconstrained model. It is often used when using a four-parameter logistic model.

Equivalence test: A method of determination that suitability testing is appropriate when the difference in the indicators used for evaluation fall within a certain range. If the difference is within the acceptable range for equivalence, the index can be determined to meet the requirements for the suitability testing. Regarding the responses obtained in the test and the coefficients of the regression equation, the indicators to be judged and the acceptable range for each indicator are established in advance. The judgment may be made based on the difference in the indicator itself, or on the confidence interval of the difference.

QC Sample: QC Sample is prepared from a representative lot at an appropriate concentration and used to evaluate the assay system. QC sample can be used to judge the suitability testing, as well as to monitor long-term changes in response. The use of QC sample is also useful for verifying that there are no changes in response before and after updating the lot of the reference material. It is also sometimes called product control or assay control.

References

1) US Pharmacopeia. (2023), <111> Design and analysis of biological assays.
2) US Pharmacopeia. (2023), <1032> Design and development of biological assays.
3) US Pharmacopeia. (2023), <1033> Biological assay validation.
4) US Pharmacopeia. (2023), <1034> Analysis of biological assays.
5) European Pharmacopoeia. 11.0 (2022), 5.3 Statistical analysis of results of biological assays and tests.