Pharmaceutical Research, Vol. 20, No. 11, November 2003 (© 2003) *Research Paper*

Recommendations for the Bioanalytical Method Validation of Ligand-binding Assays to Support Pharmacokinetic Assessments of Macromolecules

Binodh DeSilva,[1] Wendell Smith,[2] Russell Weiner,[3] Marian Kelley,[4,11] JoMarie Smolec,[5] Ben Lee,[6] Masood Khan,[7] Richard Tacey,[8] Howard Hill,[9] and Abbie Celniker[10]

Received July 2, 2003; accepted July 30, 2003

Purpose. With this publication a subcommittee of the AAPS Ligand Binding Assay Bioanalytical Focus Group (LBABFG) makes recommendations for the development, validation, and implementation of ligand binding assays (LBAs) that are intended to support pharmacokinetic and toxicokinetic assessments of macromolecules.

Methods. This subcommittee was comprised of 10 members representing Pharmaceutical, Biotechnology, and the contract research organization industries from the United States, Canada, and Europe. Each section of this consensus document addresses a specific analytical performance characteristic or aspect for validation of a LBA. Within each section the topics are organized by an assay's life cycle, the development phase, pre-study validation, and in-study validation. Because unique issues often accompany bioanalytical assays for macromolecules, this document should be viewed as a guide for designing and conducting the validation of ligand binding assays.

Results. Values of ±20% (25% at the lower limit of quantification [LLOQ]) are recommended as default acceptance criteria for accuracy (% relative error [RE], mean bias) and interbatch precision (%coefficient of variation [CV]). In addition, we propose as secondary criteria for method acceptance that the sum of the interbatch precision (%CV) and the absolute value of the mean bias (%RE) be less than or equal to 30%. This added criterion is recommended to help ensure that in-study runs of test samples will meet the proposed run acceptance criteria of 4-6-30. Exceptions to the proposed process and acceptance criteria are appropriate when accompanied by a sound scientific rationale.

Conclusions. In this consensus document, we attempt to make recommendations that are based on bioanalytical best practices and statistical thinking for development and validation of LBAs.

KEY WORDS: best practices; bioanalytical assay; biological matrices; consensus document; immunoassay.

HISTORICAL PERSPECTIVE

The first conference on bioanalytic methods validation to support bioavailability, bioequivalence, and pharmacokinetic studies was held more than a decade ago in Crystal City, VA (1,2). The proceedings and recommendations of that meeting essentially became a *de facto* guideline for bioanalytical methods validation within the pharmaceutical industry. The conference addressed validation of bioanalytical methods in general, but acknowledged differences between chromatographic and nonchromatographic assays, including immunoassays and microbiological-based methods. After the original Crystal City conference, bioanalytical methods validation was addressed subsequently several times at meetings and in publications (3–6). Most emphasis to date has been on validation of bioanalytical methods for conventional small molecular weight drugs, principally due to the rapid growth during the 1990s in the use of hyphenated mass-spectrometry as a routine analytical tool.

Because of the evolution in divergent analytical technologies for conventional small molecules and macromolecules and the growth in the interest of macromolecular therapeutics, the topic of bioanalytical methods validation was revisited in 2000 in 2 meetings, one focused on small molecule analytes (7) and the other focused on macromolecules (8). The macromolecule workshop addressed diverse topics and issues including quantitative immunoassays, bioassays, biomarkers, and assays for neutralizing antibodies. Additionally, the workshop report provided some default criteria for method and run acceptance of assays for macromolecules across a broad range of analytical methodologies (8). Because this meeting was the first one focused solely on bioanalytical methods validation for assays of macromolecules and it covered numerous topics, it was clear at the meeting's end that more in-depth discussions were needed to address the unique analytical aspects of ligand-binding assays (LBAs) for macromolecules and achieve a confluence of opinions for harmonization of bioanalytical methods validation. Accordingly, a group of the meeting's attendees formed the Ligand Binding Assay Bioanalytical Focus Group (LBABFG) within the American Association of Pharmaceutical Scientists (AAPS) to create a forum to address issues and promote education for bioanalysis of a broad range of analytes by LBAs. Because guidance is lacking for bioanalytical methods validation for assays of macromolecules, a subcommittee of the LBABFG undertook development of this consensus report. This document builds on recent publications for validation of immunoassays (3,8,9) and makes specific recommendations for validation of LBAs to support pharmacokinetics assessments of macromolecules, includes examples of prestudy validation data, summary tables, and description of calculations. For this document, LBA is synonymous with immunoassay and refers to any analytical method, in which quantification is based on macromolecular interactions.

---

[1] Amgen Inc, Thousand Oaks, California.

[2] Lilly Research Laboratories, Greenfield, Indiana.

[3] Bristol-Myers Squibb, Princeton, New Jersey.

[4] Johnson and Johnson Pharmaceutical, Raritan, New Jersey.

[5] Alta Analytical Laboratory, San Diego, California.

[6] Pfizer Global Research and Development, Ann Arbor, Michigan.

[7] Covance Laboratories, Chantilly, Virginia.

[8] PPD Development, Richmond, Virginia.

[9] HLS, Cambridgeshire, UK.

[10] Millennium Pharmaceuticals, Cambridge, Massachusetts.

[11] To whom correspondence should be addressed. (e-mail: mkelley@prdus.jnj.com)

ABBREVIATIONS: AAPS, American Association of Pharmaceutical Scientists; ANOVA, Analysis of variance; CSF, Cerebral spinal fluid; CV, Coefficient of variation; ELISA, enzyme-linked immunosorbent assay; HQC, high quality control; LBABFG, Ligand-binding Assay Bioanalytical Focus Group; LBA, ligand-binding assay; LLOQ, lower limit of quantification; LQC, lower quality control; MRD, minimum required dilution; MQC, mid-quality control; QC, quality control; ULOQ, upper limit of quantification; RE, relative error; RIA, radioimmunoassay; SOP, standard operating procedure.

INTRODUCTION

An assay life cycle can be categorized into 3 general phases: method development, prestudy validation, and in-study validation. During method development, an assay concept is evaluated, which will be confirmed during the prestudy validation phase, and applied during the in-study validation phase.

To ensure that an assay can be used in the quantification of a macromolecule in support of pharmacokinetics assessment and withstand regulatory scrutiny, specific assay components need to be evaluated, subsequently validated, and continually monitored. Therefore, validation is a continuing process as long as the assay is in use.

Validation is not a static process but a dynamic one. The components to be evaluated and the recommendations for each of the stages in the assay lifecycle are summarized in Table I. This paper is divided into these same components. Under each of the components more complete details are discussed for possible activities in each phase.

DOCUMENTATION

Method Development

The information generated during the development of a method should be documented in a laboratory notebook or other acceptable format. Information on the following assessments should be generated during assay development: critical assay reagent selection and stability, assay format selection (antibody[ies], diluents, plates, detection system etc), standard-curve model selection, matrix selection, specificity of the reagents, sample preparation, preliminary stability and a preliminary assessment of assay robustness. At the end of the development portion of the assay life cycle, a draft method or an assay worksheet should be generated to be referred to during the prestudy validation.

Prestudy Validation

A validation plan should be written before the initiation of the prestudy validation experiments. Alternatively, reference to an appropriate standard operating procedure (SOP) can be made to ensure that a documented outline exists for the experiments required for prestudy validation. This plan can be a stand-alone document or can be contained in a laboratory notebook or some comparable format. The documentation should include a description of the intended use of the method under consideration and a summary of the performance parameters to be validated that should include, but may not be limited to standard curve, precision and accuracy, range of quantification, specificity and selectivity, stability, dilutional linearity, robustness, batch size, and run acceptance criteria. The plan should include a summary of the proposed experiments and the target acceptance criteria for each performance parameter studied.

After completion of the validation experiments, a comprehensive report should be written. The format of the report may be dictated by internal policies of the laboratory; however, such reports should summarize the assay performance data and deviations from the method SOP or validation plan and any other relevant information related to the conditions under which the assay can be used without infringing the acceptance criteria.

Table I. Summary of Validation Assessment Parameters

Performance parameter	Methods development	Prestudy validation	In-study validation
Assay Reagent Selection/Stability/Assay Format/Batch size	Identify; Establish	Confirm	Monitor; Lot changes require confirmation of performance
Specificity and Selectivity	Establish	Confirm	Confirm in diseased states as available
Matrix Selection/Sample Preparation/Minimum Required Dilution	Establish	Confirm; For modified matrices, QCs must be prepared in relevant neat matrix	Monitor; With changes to lot of matrix, comparability must be demonstrated
Standard Calibrators and Standard Curves	Select Model	Confirm	Monitor
Precision and Accuracy	Evaluate imprecision and bias	Establish imprecision and bias	Total Error (4-6-30)
Range of Quantification	Evaluate	Establish	Apply
Sample Stability	Initiate	Establish	Ongoing Assessment and Extension
Dilutional Linearity	Evaluate	Establish	Monitor and establish dilutions not covered in Prestudy validation
Parallelism	N/A	Investigate where possible	Establish with incurred sample
Robustness/Ruggedness	Evaluate	Establish	Monitor
Partial Validation/Method Transfer/Cross Validation	N/A	N/A	Apply as appropriate
Run Acceptance Criteria	N/A	Runs accepted based on Standard Curve acceptance criteria	Standard Curve and QC acceptance criteria Apply 4-6-30 rule

N/A = Not applicable.

In-Study Validation

Cumulative standard curve and quality-control (QC) data tables containing appropriate statistical parameters should be generated and included with the study sample values in the final study report. Unlike the prestudy validation, failed runs are not included in these tables. Additional information to be included in the final report is a description of deviations that occurred during the study, a table of the samples that underwent repeat analysis and the reasons, and a table with details on all the failed runs.

ASSAY REAGENT SELECTION, STABILITY, ASSAY FORMAT, AND BATCH SIZE

The critical building blocks of LBAs are the ligand reagents, which typically will be an antibody or a pair of antibodies for immunoassays. Other reagents may include binding proteins, receptors, oligonucleotides, and peptide fragments. For simplicity, only immunoassay reagents are discussed in this paper. These reagents must be selected in a manner that allows for suitable specificity and selectivity and should have binding characteristics that allow for durable and stable antibody/antigen complex formation.

Assay formats can include but are not limited to, sandwich, competition, direct or indirect binding, inhibition, and solid phase or solution phase assays.

A typical batch is defined as a set of standard calibrators, validation samples and/or QC samples, and/or study samples that are analyzed together as a single group using the method to be validated. Ideally, the minimum number of QC samples should represent 5% of the total study samples in a given batch. For microtiter plate-based assays, a batch may comprise several individual plates, but each plate should contain an individual set of the standard calibrators, and QC samples.

Method Development

The first step in assay development is the selection of the assay format. For solid-phase assays, the selection of the solid phase must be considered for LBAs requiring the immobilization of the antibody or other proteins (eg, receptors). Plastic and glass surfaces, which are commonly used solid phases, can differ greatly and effect both the sensitivity and variability of an assay. The chemistry used to coat the solid phase must be considered if active adsorption (eg, avidin-biotin) is to be used. Passive adsorption is the most commonly used coating method, and can be affected by buffer characteristics, such as salt concentration and pH. The concentration of the coating solution, the mode of mixing (shaking vs. vortexing) the protein, temperature of coating, and the volumes used must be considered.

When selecting assay diluents, the specific characteristics of the analyte, intended matrix, and the binding entities (eg, antibodies or receptors) must be considered. For example, the addition of heavy metals or chelating agents may be required to enable the necessary confirmation for optimal binding. Additionally, the need for detergents (eg,Tween-20 or Triton-X100) or bulking proteins (eg, albumin, casein, or gelatin) must be considered to optimize assay performance.

Consideration should be given to selecting the assay detection system to provide an acceptable signal and to minimize noise. The sensitivity of the assay can be improved by using detection systems other than colorimetric. These systems include fluorimetric, chemiluminescence, radiometric, and electrochemiluminescence. Additional improvements may be realized through the use of signal amplification systems.

In addition to the individual components of the assay, the manner in which the assay is set up and run should be considered during method development. The assay configuration (eg, the placement of standards, QC samples, and study samples on a plate) should be established, and an attempt should be made to run batches the same size as is anticipated during in-study analysis (ie, same number of plates). The number of replicate wells used for standards, QC samples, and study samples should be determined based on the relationship between the assay's performance and the desired assay precision. While results can be based on a single point analysis or a replicate analysis, the number of replicate values that determine a QC sample result must be the same as the number of replicate values used to obtain a sample result.

The assessment of intrabatch and interbatch assay performance requires that multiple results be generated for each validation sample, where each result is calculated from response values in replicate wells. The number of batch runs and the number of replicate results (ie, repeated measurements) per sample in a run should be large enough to provide reliable estimates of assay performance characteristics. Recommendations on the minimum number of runs/replicates are proposed later in sections that address specific performance characteristics (eg, precision and accuracy).

Prestudy Validation

The critical assay reagents identified and optimized during the development phase should be used during validation without further changes. The critical assay reagents should be identified in the method documentation. The performance of the method with these critical assay reagents as defined in the draft method will be confirmed during the assay validation process.

The assay format (the number of plates, the placement of standards and validation samples, the number of replicates, operating conditions, etc) and the batch size that was established during method development should also be confirmed during prestudy validation.

In-Study Validation

During the life of an assay, often a critical component (eg, reagent lot) may need to be changed. Changes to the validated components require a partial validation to demonstrate comparable performance. Extension of the expiration dates of critical reagents (excluding reference standard) may be justified by the acceptable performance of the QC samples as they are routinely run in the assay, and extension of an expiration date must be carefully documented based on these data.

The assay format validated previously should be used during in-study validation. Any changes to this format should be clearly documented, and if a major change was done, a partial validation may be needed to demonstrate comparable performance.

REFERENCE MATERIAL

The source of materials to be used for the preparation of standard calibrators, validation samples, and QC samples will vary. It is important to clearly state the source of the material and to refer to any documentation describing characteristics of that material. Whenever possible, the standards, the validation samples, and QC samples should be prepared from separate aliquots of the same source material. In circumstances of limited availability of reference standard material (eg, drug interaction studies where the pure standard is not available except from limiting source of commercial kits), standards and QC samples can be prepared from single aliquots after checking the comparability between lots or other commercial sources. Lot numbers, batch numbers, and supporting documentation should be carefully monitored.

SPECIFICITY AND SELECTIVITY

The specificity of an antibody refers to its ability to bind the antigen of interest. Ideally, the antibody used would be specific for the target analyte, without cross-reactivity with variant forms of the analyte or other structurally related compounds that may be present in the sample. Specificity is sometimes related to the concept of cross reactivity. If an antibody is highly specific, it has low cross reactivity.

Selectivity, a concept related to specificity, is the ability of an assay to measure the analyte of interest in the presence of other constituents in the sample. Generally, LBAs measure analytes directly in the biological matrix without sample pretreatment (eg, extraction), and could be susceptible to nonspecific responses caused by reagent cross reactivity with matrix components. Specificity and selectivity evaluations verify that the assay is specific for the intended analyte and can select the analyte from a complex matrix without positive or negative interference.

Method Development

The specificity of the assay will depend on the preestablished specificity of the antibody or antibody pairs used. Antibodies may be obtained commercially or generated internally. In either case, data describing the binding characteristics of the antibody must be considered before selection.

Assay specificity is typically evaluated by assaying a sample matrix that has been spiked with increasing concentrations of variant forms of the analyte or physiochemically similar compounds, and compounds that could potentially be co administered with the analyte. In some cases, specificity can be evaluated using a sample matrix that contains concentrations of a compound or compounds related to the analyte in quantities comparable to the *in vivo* state. In general, competition assays are more susceptible to interfering compounds than sandwich-type assays, where more than one antibody confers greater specificity.

If variant or related forms of the analyte compound are not available, the generation of cross reactivity (assay specificity) data during method development may not be possible. Retrospective assessment of specificity may be required for validated assays, as more data concerning the behavior of the analyte will be generated over time.

The evaluation of selectivity during methods development assesses the quantification of the analyte in the presence of matrix components that could potentially interfere with the antibody binding. Multiple lots (at least 10 sources) of the sample matrix type should be spiked, at or near the lower limit of quantification (LLOQ) (ie, lower than the low QC sample), and assessed for percent relative error (%RE). While selectivity problems most routinely occur at the low end of an assay range, it may be prudent also to evaluate selectivity at higher analyte concentrations. In cases where interference is concentration dependent, it is essential to establish the concentrations below which interference is possible. It may be necessary to adjust the lower level of quantification accordingly, before assay validation.

Prestudy Validation

During prestudy validation, the performance of the assay with respect to specificity and selectivity is confirmed with the most relevant compounds and matrices.

Selectivity is expressed as acceptable recovery, using the same criteria that are applied during the assessment of accuracy. The recommended target acceptance criteria for selectivity, is that acceptable recovery is obtained in at least 80% of the matrices evaluated.

In-Study Validation

Typically, acceptance criteria do not exist for specificity and selectivity during sample analysis. If the potential for assay interference is an issue, sample performance characteristics must be determined *a priori* and applied routinely. It is often the case that diseased state matrices contain components (eg, rheumatoid factor, shed soluble receptors, heterophilic interferences from autoimmune diseases, lipemic samples, hemolyzed samples) that may not be present in control matrices. We highly recommended that the selectivity and specificity experiments be repeated when relevant disease-state matrices become available. The substitution of a lot of antibody requires that the specificity and selectivity evaluations be confirmed.

MATRIX SELECTION, SAMPLE PREPARATION, AND MINIMUM REQUIRED DILUTION

Considerations regarding matrix selection are one of the key differences between the assays developed for small molecule analysis and LBAs developed for the quantification of macromolecules. Because small molecule assays often include a preassay extraction, this clean-up step is often helpful to alleviate problems from individual matrix variability. The inherent characteristics of protein analytes make it difficult and often impossible to extract samples before analysis. Assays to quantitate large molecules, therefore, are often developed to measure analyte in complex matrices without extraction. In cases where endogenous protein analytes may be present, special considerations must be made for matrix selection and for analysis of data.

Method Development

Biological samples are collected in varied matrices, including plasma, serum, urine, cerebrospinal fluid (CSF), synovial fluid, and homogenized tissue. Since the characteristics of the analyte can be affected by the methods used for sample preparation, the need for any additives at the time of collection (anticoagulants, protease inhibitors, etc), stability of the analyte during collection procedures (whole blood before separation of plasma or serum), and postcollection processing and storage conditions (temperature, vial type, etc) must be evaluated. Assay format, sample collection conditions, and other factors may influence the choice of matrix in the assay. As an example, if an automated system were to be used, serum would be the preferred matrix since plasma often contains fibrin clots that may interfere with the pipetting operation. In contrast, plasma is the preferred matrix for labile analytes due to the extended time needed for the preparation of serum and because of the presence of proteolytic enzymes. More formalized stability assessments that are required for the analyte in postcollection matrix are described later in this paper.

Spiked samples should be prepared in the same matrix as the anticipated matrix of the unknown study samples. Whenever possible, the same matrix should be used to prepare the standards (calibrators)

When no endogenous signal is detectable, simple spike recovery studies will suffice to qualify a matrix as appropriate (3). Under these conditions, the use of the nominal concentration is recommended. In contrast, when a quantifiable amount of endogenous material is present in the matrix and recovery from the endogenous material and added analyte behave in an additive manner (linear), a correction factor may be applied (eg, endogenous concentration, 4 ng/mL; spiked analyte, 6 ng/mL; additive effect, 10 ng/mL). Under these circumstances, it would be appropriate to create an assigned value by adding measured endogenous concentration to the added analyte concentrations (nominal) (Table IIA and Table IIB). Another option is to subtract the endogenous concentration, measured using either the method being validated or an orthogonal method. This method requires the analysis of a blank sample (zero spike) so that the endogenous amount is subtracted and the nominal amount of the added spikes can be calculated. When the endogenous concentration and added analyte concentrations are not linear, the 2 concentrations cannot be mathematically added, and only the measured (observed) concentration is reported (Table IIC). Alternatively, the blank matrix may be stripped of the endogenous drug before spiking.

Table IIA. Example of Effect of Additive Contribution of Endogenous Protein. In this Case, the Impact of the Endogenous Protein is Seen in the Low QC. It does not Greatly Affect the Medium or High QC.

Level	Endogenous	Target QC	Observed	% Recovery
Low	10	12	22	180
Mid	10	75	85	113
High	10	250	260	104

Table IIB. Example of Additive Contribution of Endogenous Protein when Equal Volumes of Matrix and QC are Added Together. Method: 50 μL Matrix Containing Endogenous Drug (25.7 U/L) + 50 μL Std 10 (10 U/L) = Expected Value of 17.9 U/L; (1:2 dilution into each other = 12.85 U/L + 5 U/L = 17.85 U/L).

Spike recovery sample	Endogenous (U/L)	Added (U/L)	Expected (U/L)	Observed (U/L)	% Recovery
Low	25.7	10	17.9	17.6	98.3
Med	53.3	20	36.7	34.5	94
High	48.9	80	64.5	65.4	101.6

Table IIC. Example of Nonlinear Contribution of Endogenous Protein. The Observed Concentration of Drug is not Linearly Related to Either the Amount of Endogenous Drug in "Blank" Matrix or the Added "Target" Concentration. In this Case, the Observed Concentration is the Only Value That can be Used with Confidence.

Level	Endogenous	Target QC	Observed	% Recovery
Low	10	12	2	17
Mid	10	75	60	80
High	10	250	200	80

QC = quality control

The use of stripped or alternative matrix is not typically recommended but is necessary when no other strategy for quantification can be designed for measuring endogenous analytes. In a charcoal-stripped matrix, it is difficult to ensure that other aspects of the matrix did not change. Stripping by ligand-based affinity chromatography has a potential to leach ligand, which may interfere with the assay.

Regardless of the source of the matrix interference, QC samples must be prepared using the same type of neat, unaltered matrix as the study samples. In addition to assuring the stability of the analyte, preparation of QC samples in neat matrix allows for the concurrent determination of precision and accuracy of the assay.

When the matrix is difficult to procure (eg, synovial fluid, CSF), the standard curve may be prepared in a substitute matrix. An attempt should be made to prepare at least one level of QC sample in the actual sample matrix to demonstrate the absence of matrix effects, i.e., mean bias (accuracy).

The minimum required dilution (MRD) for an assay is the minimum magnitude of dilution to which a sample must be subjected to optimize accuracy and precision in an assay run with a specified standard and sample diluent. In some cases, when the standard is prepared in 100% serum or plasma, no MRD exists, and samples can be assayed undiluted or neat. In other cases, where the endogenous material does not generate a linear signal or a background signal is observed that is not due to an endogenous version of the analyte (eg, heterophilic antibodies, rheumatoid factor) dilution of the sample may be required to establish acceptable linearity. To overcome the signal-to-noise issues, all samples must be minimally diluted by some factor that results in higher reproducibility and greater accuracy throughout the assay range.

Prestudy Validation

The matrix selected during method development will be used to prepare the standard and validation samples for prestudy validation. Once the effect of the matrix has been resolved during method development, no further validation is required; however, a discussion of the resolution should be included in the validation report. In the event that a stripped matrix is used, validation samples must be prepared in neat matrix to evaluate precision and accuracy. The MRD selected during method development will be confirmed during prestudy validation.

In-Study Validation

During in-study validation, no acceptance criteria is associated with sample preparation, other than the consideration of the need for additional dilution of samples if a pattern of interference is observed. Guidance for reanalysis of samples due to potential matrix interference should be set *a priori* and applied routinely. In a situation where the matrix lot changes, QC samples should be prepared at the same concentration during prestudy validation and comparability shown before continuing sample analysis.

STANDARD CALIBRATORS AND STANDARD CURVES

Standard calibrators are prepared by spiking known amounts of a well-characterized standard material (analyte) into an appropriate matrix to generate a concentration-response relationship (standard curve) under established assay conditions, from which the analyte concentrations in unknown samples are interpolated. Standard concentration and a regression model for fitting a curve to calibration data should be established during method development, confirmed during prestudy validation, and used during in-study sample analysis. Design and analysis recommendations for the assessment of standard curve data and a regression model are summarized in Table III. Statistical analysis procedures will be illustrated using the prestudy validation standard curve data in Table IV.

Method Development

Because standard concentrations should not be changed once assay validation has begun, a greater number of standard points and replicates should be included during method development to facilitate an early detailed investigation of the concentration-response relationship. The number of standard points and replicates should be large enough to provide a reliable assessment of all competing regression models. Standard calibrators should span the anticipated concentration range for diluted study samples with concentrations approximately evenly spaced on a logarithmic scale. A minimum of 10 non-zero standard points in duplicate is recommended for the early characterization of a concentration-response relationship fit using the 4/5-parameter logistic (4/5PL) function. The standard curve should be fit to the calibrator mean response values (Fig. 1A). The decision to weight or not weight the mean responses when fitting a curve should be supported by an evaluation of the relationship between the standard deviations of the replicate values and the mean values at different concentration levels. For example, a plot of the pooled standard deviations vs. the mean responses for the standard calibrators in Table IV reveals a linearly increasing relationship on a logarithmic scale (Fig. 2). In this case, a weighted fit is generally more appropriate than an unweighted one. The information on replicate variation is useful when considering criteria for deleting a replicate value.

It is recommended that standard curve concentration-response data from a minimum of 3 independent batch runs be analyzed when establishing a calibration model. Duplicate curves should be included in each run to permit an evaluation of the intrabatch standard curve repeatability. The appropriateness of a model is judged by analysis of the RE for back-calculated standard points, where the RE for a point is obtained by subtracting the nominal concentration from the calculated concentration and dividing the difference by the nominal concentration. As a rule, the absolute RE for each back-calculated standard point should be ≤20% for most points within a curve. For example, the absolute REs of back-calculated concentrations from a weighted 5PL regression curve fit to the mean absorbance values for batch 6 data in Table IV are all <10% (Table V and Fig. 1B). Similarly, the absolute REs for the other 5 batches are also consistently <10% (except for a few points in batch 1). For a model to be considered acceptable, the accumulated back-calculated values from all curves should have an absolute mean RE of ≤10% and a coefficient of variation (CV) of ≤15% (see Table V) for all concentrations within the anticipated range. A trend in the mean RE across calibrator concentrations, i.e., calculated concentrations consistently above or consistently below the nominal concentrations, is evidence for a lack-of-fit that can invalidate a model. Lack-of-fit may be due to the

Table III. Standard Curve Assessment Criteria. "Anchor" Points are not Evaluated in the Assessment.

Assessment topic	Method development	Prestudy validation	In-study validation
Design			
1. Number of batch runs	≥3	≥6	1
2. Number of replicate curves/run	≥2	≥1	≥1
3. Number of standard points/curve	≥10	≥6	≥6
4. Number of replicate wells/point	≥2	≥2	≥2
5. Spacing of standard points	Approximately equally spaced on a logarithmic scale.		
Analysis			
6. Compute performance statistics			
a. Intracurve	%CV of replicate wells; %RE of back-calculated standard point		
b. Intercurve	Mean %RE and %CV of back-calculated standard		N/A
7. Compare statistics to target limits			
a. %CV of replicate wells (intracurve)	Specify criteria for measured responses or back calculated concentration values		
b. %RE of standard point (intracurve)	±20	±20 (±25 at LLOQ)	±20 (±25 at LLOQ)
c. Intercurve mean %RE for each calibrator	±10	≤15 (≤20 at LLOQ)	N/A
d. Intercurve %CV for each calibrator	≤15	≤15 (≤20 at LLOQ)	N/A
Acceptance Criteria			
Standard curve	Limits 7b satisfied for ≥75% of standard points.		
Regression model	Limits 7c and 7d satisfied for all points within prospective study sample concentration range		N/A

N/A = Not applicable; LLOQ = lower limit of quantification; %CV = percent coefficient of variation; %RE = percent relative error.

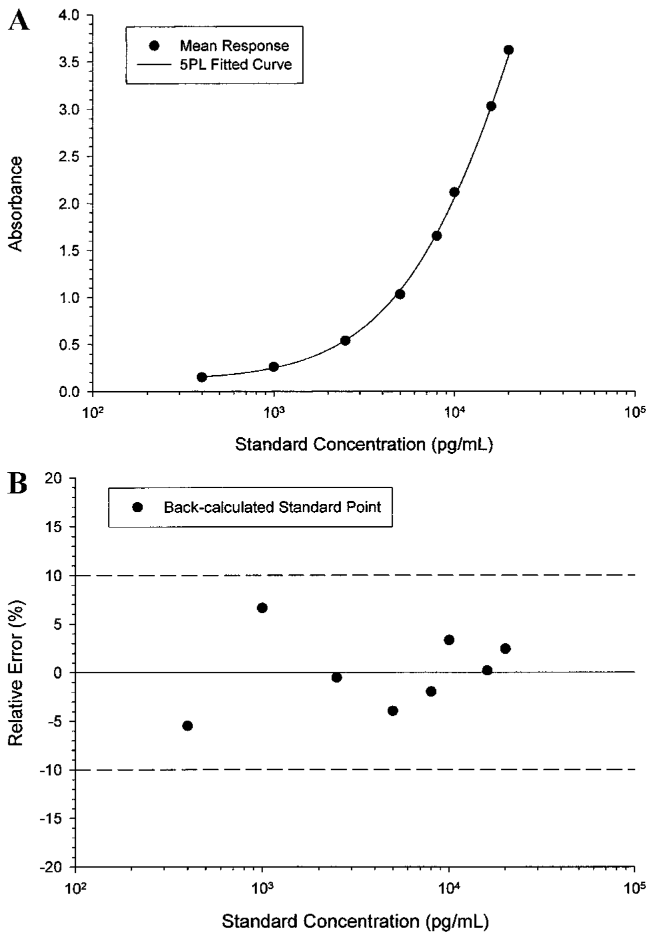

Fig. 1. A, Illustration of a five-parameter logistic (5PL) weighted nonlinear least squares regression curve. Mean responses are for the Batch 6 duplicate values listed in Table IV. B, Illustration of the percent relative errors in back-calculated concentrations for standard points in Figure 1A. All values are within 10 percent of nominal.

choice of an inappropriate mean function (eg, using the 4PL function when the response curve is asymmetric) or an inappropriate weighting procedure. For example, the use of weighting results in better curve fits to the data in Table IV as evidenced by a reduction in the mean RE (Fig. 3A) and the CV (Fig. 3B) at lower concentrations with minimal loss of accuracy or precision at higher concentrations.

Table IV. Standard Curve Response Values From a Prestudy Validation

Batch run	Nominal concentration (pg/mL)							
	400	1000	2500	5000	8000	10000	16000	20000
1	0.124	0.237	0.435	0.858	1.155	1.664	2.622	3.293
	0.131	0.213	0.479	0.938	1.313	1.732	2.834	3.519
2	0.137	0.231	0.488	0.910	1.416	1.732	2.730	3.140
	0.137	0.242	0.504	0.946	1.512	1.910	2.719	3.231
3	0.163	0.277	0.540	0.981	1.538	1.938	2.935	3.494
	0.142	0.258	0.541	1.025	1.634	2.043	3.065	3.507
4	0.128	X	0.580	1.017	1.538	1.836	2.685	2.974
	0.133	0.267	0.585	1.083	1.589	1.960	2.652	2.950
5	0.161	0.275	0.586	1.108	1.721	2.152	3.135	3.640
	0.158	0.292	0.628	1.208	1.842	2.303	3.303	3.900
6	0.149	0.254	0.511	1.004	1.574	2.043	2.920	3.400
	0.154	0.270	0.572	1.061	1.729	2.188	3.136	3.839

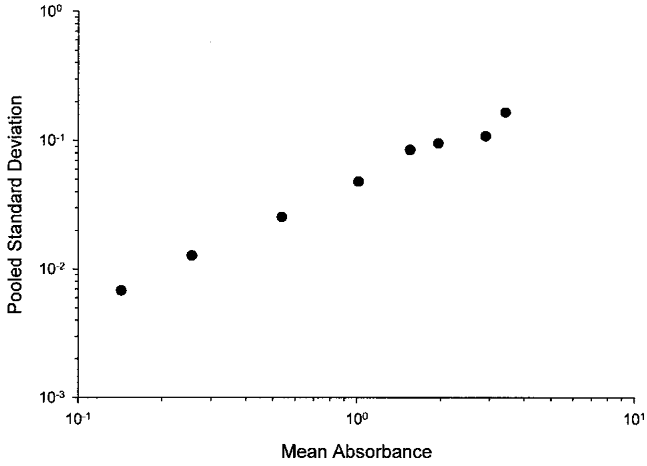

Fig. 2. Illustration of the relationship between the pooled intrabatch standard deviation and the cumulative mean response for standard curve absorbance values listed in Table IV. Each point represents the mean and standard deviation for a concentration level calculated by an analysis of variance (ANOVA). The greater variability at higher absorbance levels suggests that weighting may be appropriate when fitting standard curves.

Prestudy Validation

Standard calibrators for prestudy validation should be prepared based on performance information acquired during method development. Usually, a minimum of 6 non-zero calibrators in duplicate within the anticipated range should be included for a concentration-response relationship fit using the 4/5PL function. Standard calibrators should be spaced approximately evenly on a logarithmic scale within the anticipated range. Anchor points or calibrators that are outside the range of quantification can be included to facilitate curve fitting (8).

The regression model established during method development should be confirmed in a minimum of 6 independent prestudy validation runs, typically in the same runs in which method precision and accuracy are assessed. For the curve within a run to be acceptable, the %RE of the back-calculated value for at least 75% of the standard points, not including

Table V. Percent Relative Errors (%RE) of Back-Calculated Standard Concentrations for the Mean Response of Data Values in Table IV

Batch run	Nominal concentration (pg/mL)							
	400	1000	2500	5000	8000	10000	16000	20000
1	−11.5	11.5	2.2	0.8	−13.1	−2.5	6.6	15.2
2	−4.7	5.5	0.7	−3.7	−2.4	−0.0	3.0	2.4
3	−7.7	8.8	0.2	−4.7	−3.1	0.1	4.0	3.1
4	−1.3	3.0	−0.6	−3.2	−1.6	1.7	4.4	−1.2
5	−2.7	3.7	−0.3	−2.1	−2.3	1.3	1.4	1.6
6	−5.5	6.6	−0.5	−3.9	−2.0	3.3	0.2	2.4
Mean								
%RE	−5.6	6.5	0.3	−2.8	−4.1	0.7	3.3	3.9
%CV	3.6	3.2	1.1	2.0	4.5	2.0	2.3	5.7

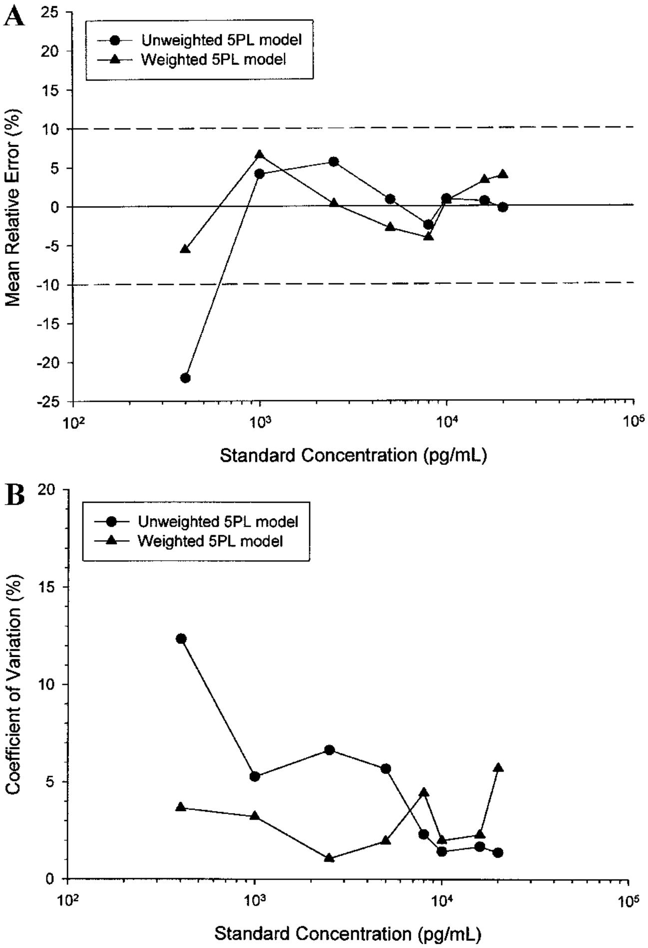

Fig. 3. A, Comparison of the cumulative mean percent relative errors in back-calculated concentrations from unweighted and weighted five-parameter logistic (5PL) regression curves fit to the absorbance data in Table IV. Relative errors for the weighted regression model are summarized Table V. The weighted model results in an acceptable fit for the entire standard concentration range while the unweighted model is unacceptable below 1000 pg/mL. B, Comparison of the cumulative percent coefficients of variation for back-calculated concentrations from unweighted and weighted five-parameter logistic (5PL) regression curves fit to the absorbance data in Table IV. Calculated values for the weighted regression model are summarized in Table V. The weighted model results in smaller interbatch coefficients of variation at lower concentration levels.

anchor points, should be within 20% of the nominal concentration, except at the LLOQ where the value should be within 25%. At the end of validation, the cumulative mean %RE and percent CV from all runs should be calculated for each calibrator. The regression model is acceptable if both the RE and CV are $\leq$15% for each standard calibrator, not including anchor points, except at the LLOQ where both should be $\leq$20%. Model confirmation should precede the reporting of analytical results for validation samples. A standard point may be edited from a curve due to a technical error with an assignable cause (eg, pipetting error) or through the application of *a priori* statistical criteria.

In-Study Validation

The standard curve for each in-study run should be monitored using the same criteria used during prestudy validation. A standard point may be edited from a curve using the same criteria established during prestudy validation. Editing must be independent of and completed before assessing QC sample performance. The final number of standard calibrators remaining after editing must be either $\geq$75% of the total, or a minimum of 6 besides the anchor points. If either of the high or low standard is deleted, the assay range for this particular run will be limited to the next standard point and samples out of range must be repeated.

PRECISION AND ACCURACY

Method precision and accuracy are performance characteristics that describe the magnitude of random errors (variation) and systematic error (mean bias) associated with repeated measurements of the same homogeneous (spiked) sample under specified conditions (9). During method development and prestudy validation, measurements are made over multiple assay batch runs with replicate determinations in each run. Factors known to vary between runs (eg, analyst, instrument, day) that may affect the variation in analytical results should be identified and documented. Method accuracy, intrabatch (within-run) precision, and interbatch (between-run) precision should be established preliminarily during method development and confirmed in prestudy validation. QC samples are used during in-study sample analysis to monitor assay performance characteristics.

Design and analysis recommendations for precision and accuracy assessments of LBAs for macromolecules are summarized in Table VI. Method precision and accuracy assessments require the comparison of calculated performance measures with *a priori* target limits that specify the amount of (analytical) error allowable without compromising the intended use and interpretation of test results (10). Limits for the minimum acceptable precision and accuracy should be established before or during method development and used throughout the assay life cycle.

Appropriate statistical methods should be used to calculate estimates of the intrabatch and interbatch precision and method accuracy (mean bias). An example of calculations applicable to method development and pre-study validation data is provided in Table VIIA, with associated formulas presented in Table VIIB and Appendix A. Intrabatch precision is estimated by the pooled intrabatch standard deviation (s_w) of measured concentration values from the calculated run means. Total random error, more commonly referred to as interbatch precision (or intermediate precision), can be estimated by the standard deviation ($s_t = \sqrt{MS_t}$) of all measured concentration values from the cumulative mean of all batches; however, the latter standard deviation is known to slightly underestimate the true interbatch imprecision. A more accurate value (s_{IP}) is calculated in Table VIIA by the method of analysis of variance (ANOVA).

Method accuracy, expressed as a %RE, is determined by the percent deviation of the weighted sample mean ($\bar{z}_w$) from the sample nominal reference value μ_T (equal to 50 ng/mL; Table VIIA). The weighted mean and sample overall mean $\bar{z}_{..}$ are equal when the number of replicates is the same for all batches. Method precision is expressed as a percent CV by dividing the calculated standard deviation by the sample nominal value. In some applications, for example when endogenous compound exists in the matrix and cannot be re-

Table VI. Precision and Accuracy Assessment Criteria. Recommended Analysis and Acceptance Criteria Apply to Each Sample Concentration

Assessment topic	Method development	Prestudy validation	In-study validation
Design			
1. Number of batch runs	≥3	≥6	1
2. Number of sample concentrations	≥8	≥5	≥3
3. Concentration levels	Span calibrator range	LLOQ, LQC, MQC, HQC, ULOQ	LQC, MQC, HQC
4. Number of replicate results/batch	≥2	≥2	≥2
Analysis			
5. Compute performance statistics	See Appendix A and Tables 5 and 6 for statistical methods and example.		%RE of QC result
6. Compare statistics to target limits			
a. Bias (%RE)	±20 (25 at LLOQ)	±20 (25 at LLOQ)	—
b. Intrabatch precision (%CV)	≤20 (25 at LLOQ)	≤20 (25 at LLOQ)	
c. Interbatch precision (%CV)	≤20 (25 at LLOQ)	≤20 (25 at LLOQ)	
d. Total error (%RE)	—	—	±30
Acceptance Criteria			
Part 1: Apply limits separately	Calculated bias and precision satisfy 6a, 6b, 6c		"4-6-30 rule"
Part 2: Combined limit	|Bias %RE| + Interbatch precision %CV ≤30		

Analysis performed for each sample concentration.

moved, a nominal value is unavailable and the calculated sample mean must be substituted for the nominal value in the percent CV computation. In this case, a scientifically defensible recovery calculation must also be recommended before validation and applied to the accuracy assessment.

Method Development

The best precision and accuracy possible for a method can be predicted early in method development by the calculated CV and mean RE of calibrators. An even more reliable measure of the expected assay performance can be obtained by preparing additional sets of independently prepared spiked control samples and evaluating them in a minimum of 3 development runs. Samples should be prepared at concentrations that span the range of calibrators (eg, 6 to 9 concentrations), with at least duplicate determinations for each concentration in each run. Interpolated results for these samples will reflect sources of variability due to calibration plus other factors related to sample preparation and location. Target

Table VIIA. Precision and Accuracy Numerical Example. Replicate Results are Analytical Data from an Immunoassay for a Therapeutic Protein. Statistics were Calculated in a Excel Spreadsheet by an Analysis of Variance (ANOVA). Symbolic Notation for all Data Values are Listed in Table VIIB with Formulae Defined in Appendix A.

		Replicate results			Intrabatch (within-run) statistics					
Sample	Batch run	1	2	3	n	Mean	SD	%CV	%RE	Ancillary statistics
QC 4 50 (ng/mL)	1	47.6	48.1	52.2	3	49.3	2.52	5.0	−1.4	MS_w = 9.320
	2	42.0	41.4	43.7	3	42.4	1.19	2.4	−15.3	MS_b = 59.444
	3	72.4X	53.1	45.8	2	49.5	5.16	10.3	−1.1	MS_t = 24.984
	4	53.4	55.3	54.5	3	54.4	0.95	1.9	8.8	s_t = 4.998
	5	45.6	42.6	51.5	3	46.6	4.53	9.1	−6.9	s_b = 4.213
	6	46.5	42.3	40.8	3	43.2	2.95	5.9	−13.6	p = 6
	Intrabatch (within-run) statistics (Pooled):				2.88	47.4	3.05	6.1	−5.1	
	Interbatch (between-run) statistics (ANOVA):				17	47.5	5.20	10.4	−5.0	

X—Analytical error, value omitted from computations.

Table VIIB. Symbolic Notation for Numerical Example in Table VIIA. All Symbols are Defined in Appendix A.

		Replicate results			Intrabatch (within-run) statistics					
Sample	Batch run	1st	2nd	3rd	n	Mean	SD	%CV	%RE	Ancillary statistics
QC# μ_T (units)	1	z_{11}	z_{12}	z_{13}	n_1	$\bar{z}_{1.}$	s_1	$100(s_1/\mu_T)$	$100(\bar{z}_{1.} - \mu_T)/\mu_T$	MS_w
	2	z_{21}	z_{22}	z_{23}	n_2	$\bar{z}_{2.}$	s_2	$100(s_2/\mu_T)$	$100(\bar{z}_{2.} - \mu_T)/\mu_T$	MS_b
	3	z_{31}	z_{32}	z_{33}	n_3	$\bar{z}_{3.}$	s_3	$100(s_3/\mu_T)$	$100(\bar{z}_{3.} - \mu_T)/\mu_T$	MS_t
	4	z_{41}	z_{42}	z_{43}	n_4	$\bar{z}_{4.}$	s_4	$100(s_4/\mu_T)$	$100(\bar{z}_{4.} - \mu_T)/\mu_T$	s_t
	5	z_{51}	z_{52}	z_{53}	n_5	$\bar{z}_{5.}$	s_5	$100(s_5/\mu_T)$	$100(\bar{z}_{5.} - \mu_T)/\mu_T$	s_b
	6	z_{61}	z_{62}	z_{63}	n_6	$\bar{z}_{6.}$	s_6	$100(s_6/\mu_T)$	$100(\bar{z}_{6.} - \mu_T)/\mu_T$	p
	Intrabatch (within-run) statistics (Pooled):				$\bar{n}$	$\bar{z}_{..}$	s_w	$100(s_w/\mu_T)$	$100(\bar{z}_{..} - \mu_T)/\mu_T$	
	Interbatch (between-run) statistics (ANOVA):				N	$\bar{z}_w$	s_{IP}	$100(s_{IP}/\mu_T)$	$100(\bar{z}_w - \mu_T)/\mu_T$	

limits of 20% (25% at LLOQ) recommended in Table VI for the cumulative %CV and absolute mean RE (mean bias) for each concentration are the same as those proposed for prestudy validation assessments of precision and accuracy (3).

Prestudy Validation

Method precision and accuracy are confirmed during prestudy validation by the analysis of validation samples, prepared in a matrix equivalent to that of anticipated unknown sample matrices (3), spiked at 5 or more concentrations (anticipated LLOQ, less than 3 times LLOQ, mid, high, and anticipated upper limit of quantification [ULOQ]). It is recommended that at least 2 independent determinations per run (replicates) be made for each sample in a minimum of 6 runs. For each validation sample, the repeated measurements from all runs should be analyzed together using an appropriate statistical method (see Table VII and Appendix A for example and formulas).

For a method to be considered acceptable, it is recommended that the interbatch precision (%CV) and the absolute mean bias (%RE) both be ≤20% (25% at LLOQ). In addition, it is recommended that the method total error (sum of the %CV and absolute %RE) be ≤30% (40% at the LLOQ) to be consistent with the in-study validation acceptance criteria.

In-Study Validation

Precision and accuracy for each in-study run is monitored by the evaluation of QC sample results. Run acceptance criteria that have been embraced for both chromatographic and LBAs require at least two thirds of all QC results for a run to be within a specific percent (eg, 15%, 20%, 25%, or 30%) of the corresponding nominal reference values, with at least 50% of the results within the specified limit for each QC sample. For assays of conventional small molecule drugs, a 4-6-15 rule has been adopted (7). In contrast, a 4-6-30 rule was proposed for LBAs of macromolecules at the March 2000 AAPS workshop (8). We recommend adoption of the proposed 4-6-30 rule.

The acceptance criterion for precision and accuracy described in the prestudy validation section was recommended because it is computationally straightforward and reasonably consistent with the desired in-study 4-6-30 rule. By definition, the 4-6-30 criteria are based solely on the deviation of individual results from their nominal values and not on a calculated mean or standard deviation. Since the deviation of an analytical result from a nominal value includes components of both random and systematic errors, it is therefore a measure of the total error (Fig. 4). The added prestudy constraint that the sum of the %CV and absolute %RE be ≤30% prevents the acceptance of an assay that has high levels of both imprecision and bias (eg, near 20%). Such an assay would be expected often to fail the in-study 4-6-30 criteria. Other statistical methods that ensure consistency between prestudy and in-study acceptance criteria are also acceptable (3,11).

RANGE OF QUANTIFICATION

For immunoassays and other LBAs, the range of quantification should be based on the lowest (LLOQ) and highest (ULOQ) validation samples that meet the target precision and accuracy criteria, rather than on the performance of the standard calibrators. Since validation samples used to define the range of quantification are prepared in undiluted sample matrix, they may be subjected to the minimal required dilution before analysis. In those cases where it is required to use a MRD it is acceptable to define the range of quantification as the standard concentration values in either neat matrix or as the range of standard concentration values obtained after applying the MRD. As an example, a standard curve of 10 to 100 ng/mL in neat matrix, is equivalent to a standard curve range of 1 to 10 ng/mL with an MRD of 10 (i.e., 10% matrix).

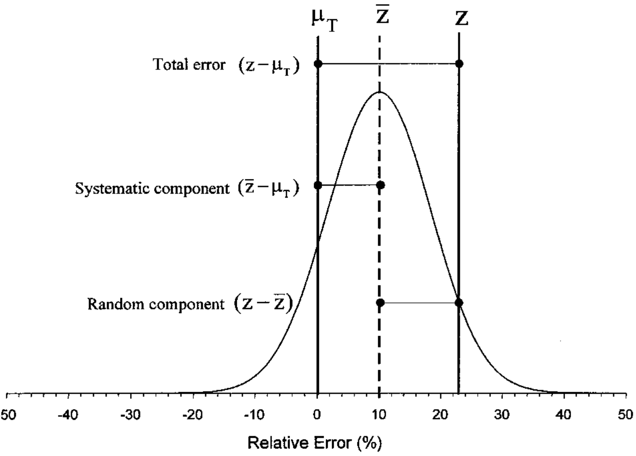

Fig. 4. Illustration of the total error for a measured analytical result (z), defined as the deviation of the result from its nominal "true" value (μ_T). Errors from repeated measurements of a homogeneous sample are generally assumed to follow a bell-shaped normal distribution. For comparison purposes, errors are often expressed as a percent relative error (see the figure scale) by dividing a deviation by the nominal value and multiplying by 100. Note that total error is equal to the sum of a systematic error component, estimated by the deviation of the calculated analytical mean ($\bar{z}$) from the nominal value, plus a random error component estimated by the deviation of a result from the analytical mean.

Method Development

Early in development, back-calculated standard values may be used to generate a preliminary estimate of the range of quantification. Later, spiked samples are used to refine the estimated range. At this stage, it is useful to include more points in the areas of the anticipated LLOQ and ULOQ. A plot of the precision profile can be helpful in assessing the prospective limits of quantification (Fig. 5).

Prestudy Validation

The reportable range of the assay should be established using results from the spiked validation samples that were prepared for the precision and accuracy assessment. Standard-curve calibrators should include concentrations that span the anticipated LLOQ and ULOQ. The LLOQ and ULOQ are determined by the lowest and highest validation samples for which the precision (interbatch %CV) and accuracy (absolute %RE) are both ≤20% (25% at the LLOQ) and the sum of the 2 is ≤30%.

In-Study Validation

The range of quantification established during prestudy validation is the range into which samples must be diluted if

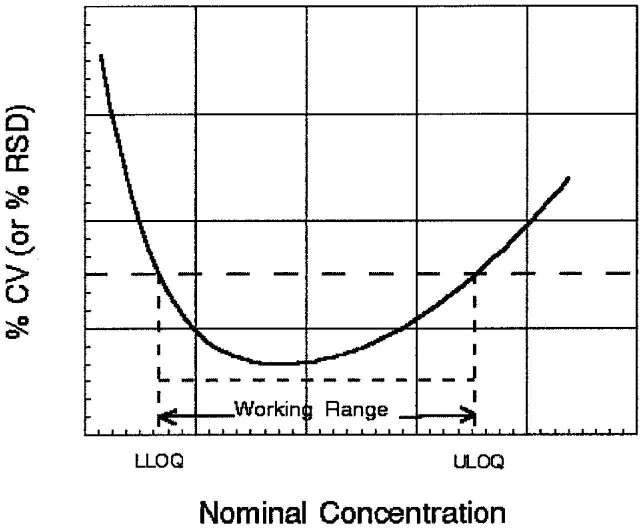

Fig. 5. Typical precision

necessary. Samples that are above the ULOQ must be reassayed at a greater dilution. Samples already at the minimal required dilution, and below the LLOQ must be reported as <LLOQ. During sample analysis, the LLOQ for a run must be revised upward if editing of the standard curve results in no calibrator at or below the validated LLOQ. In this case, the LLOQ is increased to the lowest remaining standard concentration.

SAMPLE STABILITY

Experiments that demonstrate the acceptable stability of analyte in the sample matrix must be included in the prestudy validation assessment. Stability experiments should mimic, as best possible, the conditions under which study samples will be collected, stored, and processed. Assessments are most often conducted by adding analyte to whole blood as well as the matrices that result after whole blood processing (plasma and/or serum). Processing and storage conditions assessed usually include bench-top stability, short-term and long-term storage, and the stability to multiple freeze-thaw cycles. The physiochemical properties of the analyte should be considered when handling conditions are being established. Stability of stock standard solutions under conditions of storage must be established as well.

Method Development

Stability samples must be prepared in the same matrix in which study samples are to be collected. In cases where stripped or altered matrix is to be used for preparation of study calibration standards and QC samples, the stability samples must be prepared in the unaltered matrix. The preparation of stability samples during assay development may greatly facilitate the gathering of long-term stability data during prestudy validation. Early preparation of stability samples may give one a head start on the establishment of the long-term sample stability, and should be prepared in a suitable laboratory environment accompanied by appropriate documentation. The risks associated with this approach include the chances of a matrix change as well as the dependence on the assigned nominal concentration as the reference value.

Stability assessments may be conducted during the development of an assay for elements of sample handling, including, but not limited to, evaluation of matrix stability at room temperature, 2° to 8°C, and through freeze-thaw cycles, defining how samples will be handled through out the life cycle of the assay.

Prestudy Validation

Formal stability evaluation must be conducted with an established assay during prestudy validation. The stability samples must be prepared by spiking the analyte of interest into the same matrix as the study samples at a high and a low concentration. These high and low values may be the same concentrations as the high and low QC samples. It is recommended that the stability samples be run with the same number of replicates that are used for the QC samples.

Bench top stability requires that samples be handled in a manner that mimics the processing at both the study and analytical sites, and should be done at room temperature (minimum 2 h) and at refrigerator temperature (2° to 8°C) (minimum 24 h) (12).

The stability of an analyte in whole blood is assessed by spiking analyte into freshly collected whole blood. As an example of evaluating the stability of analyte in whole blood, a whole blood sample may be spiked with analyte, and incubated for up to 2 h and processed to obtain plasma or serum at several intervals. The processed samples will be evaluated for stability by monitoring trends in recovery (12).

The freeze-thaw stability evaluation should take into account the anticipated freeze-thaw cycles during routine analysis. A standard approach is 3 freeze-thaw cycles with no less than 12 h between thaws. The rate of freezing and thawing and frozen storage temperature should mimic the manner in which samples will be handled as they are being thawed before assaying.

Long-term stability must take into account storage at the site and storage at the testing facility. It is essential to show that samples are stable throughout the lifetime of a study including post in-study sample analysis. The interval for testing depends on the needs of the study. For very long-term studies, stability analysis may need to be conducted at intervals more frequent than the sample analysis to insure that samples can be batched until the end of a study. The need to conduct studies on samples stored at both −20°C and −70° to −80°C may depend on the duration of storage at −20°C. If samples are to be frozen at −20°C, followed by storage at −80°C, the stability samples should be prepared in the same manner, with the time at −20°C modeled as well as can be predicted. Long-term stability assessments often continue during in-study validation.

A freshly prepared standard calibrator curve and QC samples that either are within the acceptable expiration or freshly prepared should be used as the reference for comparison of the stability samples. The stability evaluations, with the exception of the whole blood stability, will be acceptable using the same acceptance criteria applied for accuracy and precision of QC samples. If the measured sample value is within the acceptance criteria for accuracy, then the sample is considered stable even if a trend in the stability is observed. Alternative assessments may be applied, such as the use of confidence intervals (13). In this case when the observed concentration or response of the stability sample is beyond the lower confidence interval, the sample is no longer considered valid.

In-Study Validation

Typically, stability assessments continue during the in-study validation period. If sample handling and or storage conditions of the study samples were changed, additional stability evaluations must be done to reflect the new conditions. If samples were inadvertently stored at a different temperature, stability studies at that temperature should be done to verify stability before sample analysis. The validation report may be updated to include these data. When a stability time point assessment fails, as long as there is an *a priori* plan for confirming the trend, sample analysis can proceed within the time limits bound by the time of the sample failure, but then can be extended if reanalysis of the stability time point refutes the trend.

DILUTIONAL LINEARITY

Due to the nature or format of many immunoassays, the range of quantification (LLOQ to ULOQ) of the standard curve may be very narrow, sometimes <1 order of magnitude. For this reason it is necessary to show that the analyte of interest, when present in concentrations above the range of quantification (above ULOQ), can be diluted to bring the analyte concentrations into the validated range for analysis by the method. An additional reason for conducting dilutional experiments is for the identification of a possible prozone or "hook effect" (see Fig. 6 for example of signal suppression caused by high concentrations of analyte) (14).

Dilutional linearity should not be confused with parallelism. Parallelism must be assessed using incurred samples and dilutional linearity can be assessed using spiked QC samples. If dilutional linearity is demonstrated during the prestudy validation, it may preclude the need to run QC samples in a dilution series during in-study validation.

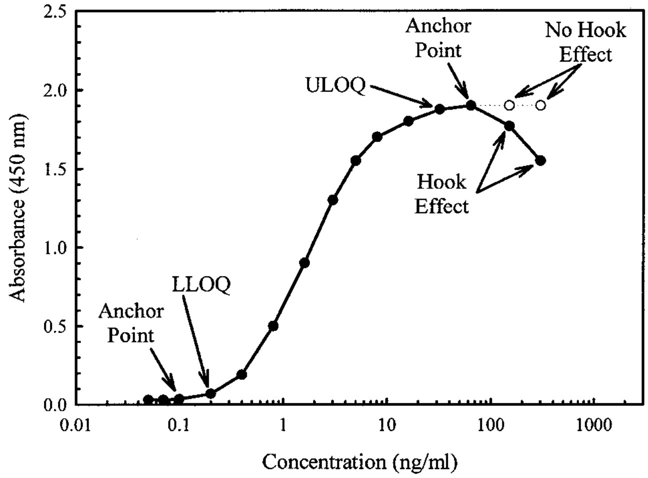

Fig. 6. Demonstration of Hook (Prozone) Effect. Typical sigmoidal dose-response curve for a two-site EIA. (•), including the high dose hook effect. Specifically, higher concentrations of analyte with a resulting lower than expected response. If there were no hook effect, as shown by the open circles (○), higher concentrations of analyte would result in a >ULOQ response. Without a hook effect, the quantification range of the curve is between the LLOQ and ULOQ. The anchor points beyond the LLOQ and ULOQ are used only for curve fitting.

Method Development

Dilutional linearity should be evaluated on spikes that have been made into the sample matrix and then are diluted into the assay matrix. The matrix can comprise individual or a pool of individual samples. The choice of pooled vs. individual samples depends on whether interference from substances, such as heterophilic antibodies or binding proteins, may be of concern. Dilutions should be made such that several dilutions fall on the standard curve.

Evaluation of dilutional linearity should be done with a spike made 100- to 1000-fold greater than the ULOQ. When this evaluation is not feasible, efforts should be made to make the concentration as high as possible. The dilutional samples that are evaluated should be designed to include a dilution above the ULOQ (to evaluate hook effect) as well as values in the upper, middle, and lower parts of the curve (to evaluate dilutional linearity). Generally, it is common practice that an individual dilution step not exceed 1:100.

When there is a lack of dilutional linearity an appropriate strategy for analysis of high concentration samples must be established. The use of a MRD or a plateau value established before reporting may accommodate this need. When linearity cannot be achieved, a strategy for reporting data (i.e., at the greatest dilution within the assay range) must be established.

Prestudy Validation

The dilutions evaluated in method development should be confirmed during prestudy validation. The back-calculated concentration for each individual dilution and the cumulative precision of the final concentrations across all the dilutions should be made. The back-calculated concentration for each diluted sample should be within 20% of the nominal or expected value. The precision of the cumulative back calculated concentration should be ≤20%. Theoretically, the dilutional linearity samples prepared 1000 times ULOQ should yield a back-calculated value of >ULOQ; however, if the back-calculated value is within the range of quantification, the hook effect may be present.

In-Study Validation

During prestudy validation, the full range of dilutions for study samples is usually covered. When a study sample needs to be diluted at a concentration exceeding that assessed during prestudy validation, dilutional linearity should be repeated including this concentration. Alternatively, a dilutional QC sample can be included in the assay.

PARALLELISM

Parallelism is a performance characteristic that is typically evaluated during in-study validation. It is conceptually similar to dilutional linearity except that it is assessed with multiple dilutions of actual study samples or samples that represent the same matrix and analyte combination as those that will be generated during a study.

Method Development

Parallelism is not typically evaluated during method development, but rather dilutional linearity is used as a first-pass estimate of parallelism.

Prestudy Validation

When an assay is being validated for a preclinical study, it is sometimes possible to obtain samples from pilot studies where animals have been exposed to high doses of the analyte. When samples of this type are available, the assessment of parallelism can be done during prestudy validation. Additionally, when an assay is being validated to replace another assay, and samples from studies conducted with the same drug substance (active pharmaceutical ingredient) are available, parallelism can also be evaluated during prestudy validation.

In-Study Validation

Parallelism can be assessed using C_{max} samples from a given study. A common approach is to pool several C_{max} samples to create a parallelism validation sample. Demonstrating parallelism of the pooled sample eliminates the generation of multiple values for individual study samples. The degree of non-parallelism that is acceptable for a method depends on its intended application (15). As a target, it is recommended that the relative standard deviation (%CV) between samples in a dilution series be $\leq$30%. The procedure for reporting a result when a sample does not dilute linearly (i.e., non-parallel) should be defined *a priori.*

ROBUSTNESS/RUGGEDNESS

The main focus of any robustness/ruggedness analysis is to address the question of whether the assay will perform well under real life changes in standard laboratory situations. While there may be considerable debate regarding the absolute defined differences between assay robustness and assay ruggedness, it is clear that both parameters are indicators of assay reproducibility under varied conditions. The 2 parameters are very closely related and therefore are being separated only to allow for more clarity in defining how to assess both at different stages in the assay development and validation life cycle.

The robustness of the assay is determined by its consistency when changes are implemented that may impact the assay. For this reason, those changes must be considered, tested, and documented. Changes that have an impact on the assay must be identified in the method or method SOP (6). Elements in an immunoassay that could impact its consistency include incubation temperatures, light exposure (ELISA), and matrix (plasma, serum, CSF).

The ruggedness of an assay is determined by its consistency when routine changes are implemented resulting in different operational conditions (4). For instance, circumstances such as changes of the analysts; different instruments; batch size; and the day, time, or other environment factors should not greatly impact the consistency of the assay.

Method Development

During development, the types of run variables that should be assessed include, but are not limited to incubation times (for all steps of the assay), incubation temperatures (for all steps), different analysts and the instruments (including pipettes, liquid handling systems, washers, and readers) that will be used to conduct the assay. A subset of these variables may be reassessed during prestudy validation, but it is important to make sure that they are evaluated before finalization of the assay method so that validation is conducted within the limits set for these parameters.

Prestudy Validation

During prestudy evaluation of robustness and ruggedness of the method, an attempt should be made to evaluate the variety of conditions that may reflect the execution and performance of the method during the in-study phase. As an example, robustness assessment could include incubation time tolerances, while ruggedness assessment could include changes in analysts and batch size. Most robustness and ruggedness evaluations are empirical in nature; however, more formal evaluations can also be used (16).

In-Study Validation

The ongoing monitoring of QC performance and intraassay and interassay precision evaluations at the end of studies can provide information of the robustness and ruggedness of an assay conducted under different conditions. For example, include incubation times that vary by about 15% to accommodate such lapses in every day sample analysis (2 h ±15 min).

PARTIAL VALIDATION, METHOD TRANSFER, AND CROSS VALIDATION

Validations can fall into 3 broad categories, full, partial, and cross. A full validation is done for any new methods as described in this paper, and involves method development, prestudy, and in-study validation. Full validations are required for change in species (eg, rat to mouse), change in matrix within a species (eg, rat serum to rat urine).

Partial Validation

A partial validation is conducted where method changes are considered to be minor in nature and this may include method transfer, changes to anticoagulant (eg, EDTA, heparin, citrate), change in method (especially pivotal reagents such as the primary antibody or secondary antibody), sample processing changes (how fast a clot needs to be spun, collection vessels, storage condition), sample volumes, extension of the concentration range, selectivity issues (concomitant medication), conversion of a manual to an automated method, qualification of an analyst, etc. Partial validations can range from a single intraassay accuracy and precision run to a nearly full validation. Changes to lots of reagents or sample-processing changes may only require 1 run. In contrast transferring an analytical method may require substantially more experimentation.

Method Transfer

Method transfer is the situation where the method is established in one laboratory (sending laboratory) and transferred to another laboratory (receiving laboratory) and requires at least a partial validation.

In addition to the required documentation (eg, method description, validation report, certificate of analysis) the sending laboratory should provide information on those factors that may affect the ruggedness of the assay (eg, identifying pivotal reagents and material). The method transfer requires a plan or protocol that defines the process (eg, experiments to be conducted) and the acceptance criteria.

Once the transferred method is validated, an ideal scenario is to have both the sending and receiving laboratories analyze 30 blinded, spiked samples covering the standard curve range, and 30 pooled incurred samples. The 2 sets of data are compared by using a statistical equivalence test (17,18). Alternatively, the differences between the 2 sets of data may be compared using agreed range of acceptability.

Cross Validation

Cross validation is conducted when 2 validated bioanalytical methods are used within the same study or submission, for example, ELISA assay to BiaCore, ELISA to a liquid chromatography/mass spectrometry. It is recommended that test samples (spiked and/or pooled incurred samples) be used to cross validate the bioanalytical methods. Data should be evaluated using an appropriate predefined acceptance criteria or statistical method (17,18).

RUN ACCEPTANCE CRITERIA

Method Development

During method development, no defined run acceptance criteria should be set. Early assessments of the performance of the standard curve can be used to judge the appropriateness of the chosen reagents and assay format.

Prestudy Validation

Prestudy validation runs are accepted based on the standard curve acceptance criteria. No run acceptance criteria are applicable for prestudy validation sample assessments, i.e., no run can be rejected due to poor validation sample performance during accuracy and precision evaluation, and all data from the prestudy validation runs are reported without exceptions. In some cases, there may be assignable cause (eg, technical issues) for removal of a validation sample data point before the calculation of the cumulative mean. Exclusion is applied at the end of the validation study period and must be documented as described in the documentation section.

In-Study Validation

For each in-study run, the standard curve must satisfy criteria described in the standard-curve section; however, run acceptance is based primarily on the performance of the QC samples. For LBAs of macromolecules, the run acceptance criteria recommended in the Precision and Accuracy Section requires that at least 4 of 6 (67%) QC results must be within 30% of their nominal values, with at least 50% of the values for each QC level satisfying the 30% limit. The recommended 4-6-30 rule imposes limits simultaneously on the allowable random error (imprecision) and systematic error (mean bias). If the application of an assay requires a QC target acceptance limit different than the 30% deviation from the nominal value, then prestudy acceptance criteria for precision and accuracy should be adjusted so that the limit for the sum of the interbatch imprecision and absolute mean RE is equal to the revised QC acceptance limit.

CONCLUSION

A primary focus of an immunoassay to support pharmacokinetic studies is in its development. By thoroughly defining each component of the immunoassay in the first phase of the assay life cycle, the resulting validation plan should be concise and the validation process straightforward. Several facets of the assay defined in the late development stage, such as specificity and dilutional linearity, can appropriately be included in the final validation report.

As described in this paper, a typical validation will include at least 6 precision and accuracy assays to define the consistency of the assay. Within those assays, several parameters can be defined, including early stability, specificity, selectivity, and range of quantification. The standard curve should be defined with at least 6 non-zero points and, during the development phase, the accuracy is assessed. No assay run should be eliminated except for a true and documented analyst's error.

The validation samples define the range of the assay and no values below the LLOQ or above the ULOQ may be reported. Within the 6 validation assay runs, the validation samples are used to define the cumulative precision and accuracy. During validation, no validation sample may be eliminated to show the true profile of the assay.

During sample analysis, the validation life cycle continues. Before assessing the QC samples for acceptance the standard curve must be deemed appropriate by predetermined criteria. Only after the curve is accepted may the assessment of QC samples continue. QC sample results determine if the assay run is valid. Acceptance criteria can be based on 4–6-× rule or on total error and should be predicated on the criteria used in both the development and prestudy validation phases. Overall, the immunoassay is a highly sensitive assay that can be used to quantify protein and peptide drugs in a biologic matrix, often routinely in the pg/mL range.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Ronald Bowsher (LINCO Diagnostic Services and previously Eli Lilly) for his critical review, advice, and constructive comments on the manuscript. We would also like to thank Dr. Jean Lee (MDS Pharma Services) and Mr. Jeff Sailstad (Trimeris) for reviewing the manuscript and providing guidance throughout this process. We sincerely thank Dr MaryAnn Foote (Amgen) for her help with the manuscript.

APPENDIX A: PRECISION AND ACCURACY STATISTICS

In Table VIIA, assay results are listed for a series of 6 batch runs in which triplicate measurements were made of a spiked sample with a known nominal theoretical concentration (μ_T) equal to 50 ng/mL. One value (first replicate of batch 3) was flagged for exclusion from the statistical analysis

Table VIII. Precision and Accuracy Summary Table. The 50 ng/mL Values for N, Mean Bias, Intrabatch and Interbatch Precision are Copied From Table VIIA. Formulas for Computation of the Lower (LCL) and Upper (UCL) Confidence Limits for Mean Bias and the 90% Expectation Tolerance Interval Limits are Listed in Appendix A.

		Nominal concentration				
Characteristic	Statistic	2 (ng/mL)	10 (ng/mL)	32 (ng/mL)	50 (ng/mL)	64 (ng/mL)
# Results	N	18	18	18	17	18
Accuracy	Mean Bias (%RE)	–2.9	–7.4	–1.1	–5.0	–2.0
	LCL	–12.2	–15.6	–7.6	–14.6	–11.8
	UCL	6.4	0.9	5.3	4.7	7.8
Precision	Intrabatch (%CV)	6.9	7.5	9.6	6.1	5.5
	Interbatch (%CV)	10.5	10.0	10.0	10.4	10.4
Accuracy + Precision	$\lvert$Mean$\rvert$ + Interbatch	13.4	17.3	11.1	15.4	12.4
90% Expectation	Lower Limit (%RE)	–23.1	–26.1	–19.1	–25.5	–22.8
Tolerance Interval	Upper Limit (%RE)	17.4	11.4	16.8	15.6	18.8

because of a documented analytical error. Calculated estimates of precision and accuracy in Table VIIA and Table VIII were obtained through an Excel spreadsheet using formulae for an analysis of variance (ANOVA) based on the one-factor random effects model (19). The model assumes that the measured result (z_{ij}) from the j^{th} replicate of the i^{th} batch run is described by the equation

$$z_{ij} = \mu_M + b_i + \varepsilon_{ij} \; (i = 1, 2, \ldots\ldots p; \; j = 1, 2, \ldots\ldots n_i),$$

where p is the total number of batch runs, n_i is the number of replicates in the i^{th} batch, μ_M is the unknown true analytical mean for the method, and b_i and ε_{ij} denote random effects that explain the uncertainty in the i^{th} batch and j^{th} replicate respectively. Random effects are assumed to be normally and independently distributed with mean zero and variance components equal to σ_b^2 for between batch runs and σ_w^2 for replicates within a batch. Statistical analysis of the measured z_{ij} values leads to estimates of the unknown parameters μ_M, σ_w^2, and σ_b^2, and assay characteristics derived from them.

The following formulae were used to compute the statistics summarized in Table VIIA and Table VIII. Note that the number of replicates per batch is not required to be equal, therefore allowing for the exclusion of the flagged value equal to 72.4.

Sample Size	Number of replicates	n_i (ith batch)
Statistics	Total number of results	$N = \sum_{i=1}^{p} n_i$
	Effective no. of replicates	$\tilde{n} = \frac{1}{N} \sum_{i=1}^{p} n_i^2$
Statistics for ith batch	Batch mean	$\bar{z}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} z_{ij}$
	Batch standard deviation	$s_i = \sqrt{\frac{1}{(n_i - 1)} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{i.})^2}$
	Percent coefficient of variation	$\%CV = 100(s_i/\mu_T)$
	Percent relative error	$\%RE = 100(\bar{z}_{i.} - \mu_T)/\mu_T$
ANOVA Mean Squares	Intrabatch	$MS_w = \frac{1}{(N - p)} \sum_{i=1}^{p} (n_i - 1)s_i^2$
	Interbatch	$MS_b = \frac{1}{(p - 1)} \sum_{i=1}^{p} n_i(\bar{z}_{i.} - \bar{z}_{..})^2$
	Total	$MS_t = \frac{1}{(N - 1)} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{..})^2$
ANOVA Variance Components	Intrabatch	$s_w^2 = MS_w$
	Interbatch	$s_b^2 = \left(\frac{p - 1}{N - \tilde{n}}\right)(MS_b - MS_w)$
	Total	$s_{tot}^2 = s_w^2 + s_b^2$
"Pooled" Intrabatch Statistics	Sample overall mean	$\bar{z}_{..} = \frac{1}{N} \sum_{i=1}^{p} n_i \bar{z}_{i.}$
	Intrabatch standard deviation (Repeatability)	$s_w = \sqrt{MS_w}$ if $MS_b > MS_w$ $= \sqrt{MS_t}$ otherwise
	Percent coefficient of variation	$\%CV = 100(s_w/\mu_T)$
	Percent relative error	$\%RE = 100(\bar{z}_{..} - \mu_T)/\mu_T$

Interbatch Statistics	Weight for ith batch	$w_i = \dfrac{n_i}{s_w^2 + n_i s_b^2}$
	Weighted sample mean	$\bar{z}_w = \sum_{i=1}^{p} w_i \bar{z}_{i.} / \sum_{i=1}^{p} w_i$
	Interbatch standard deviation (Intermediate precision)	$s_{IP} = \sqrt{s_{tot}^2}$ if $MS_b > MS_w$ $= \sqrt{MS_t}$ otherwise
	Percent coefficient of variation	$\%CV = 100(s_{IP}/\mu_T)$
	Percent relative error	$\%RE = 100(\bar{z}_w - \mu_T)/\mu_T$
Interval Statistics	Estimated variance of $\bar{z}_{..}$	$\hat{var}(\bar{z}_{..}) = \dfrac{1}{N}(s_w^2 + \bar{n}\, s_b^2)$
	Estimated variance of $\bar{z}_w$	$\hat{var}(\bar{z}_w) = \left(\sum_{i=1}^{p} w_i \right)^{-1}$
	Design constant	$a = \dfrac{(p-1)}{N - \bar{n}}$
	αth percentile of Student's t-distribution with k degrees of freedom	$t_{k,\alpha}$
	αth percentile of Chi-square distribution with k degrees of freedom	$\chi_{k,\alpha}^2$
	Degrees of freedom for $\hat{var}(\bar{z}_{..})$	$\eta = \dfrac{(s_w^2 + \bar{n}\, s_b^2)^2}{\dfrac{[(1-\bar{n}\, a)s_w^2]^2}{N-p} + \dfrac{[\bar{n}\, a\, s_w^2 + \bar{n}\, s_b^2]^2}{p-1}}$
	Degrees of freedom for s_{IP}^2	$\nu = \dfrac{(s_{IP}^2)^2}{\dfrac{[(1-a)\, s_w^2]^2}{N-p} + \dfrac{[\bar{n}\, a\, s_w^2 + s_b^2]^2}{p-1}}$
	Two-sided 95% confidence limits for mean bias (%RE)	$\left(\dfrac{100}{\mu_T}\right) \cdot [(\bar{z}_w - \mu_T) \pm t_{\eta,0.975} \cdot \sqrt{\hat{var}(\bar{z}_w)}]$
	Two-sided 90% expectation tolerance interval for relative error (20)	$\left(\dfrac{100}{\mu_T}\right) \cdot [(\bar{z}_w - \mu_T) \pm t_{v,0.95} \sqrt{\hat{var}(\bar{z}_w) + s_{IP}^2}]$

REFERENCES

1. V. P. Shah, K. K. Midha, S. V. Dighe, I. J. McGiveray, J. P. Skelly, A. Yacobi. T. Laylogg, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Putman, and S. Spector. Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. *J. Pharm. Sci.* **81**:309–312 (1992).
2. V. P. Shah, K. K. Midha, S. Dighe, I. J. McGilveray, J. P. Skelly, A. Yacobi, T. Layloff, C. T. Viswanathan, C. E. Cook, R. D. McDowall, K. A. Pittman, and S. Spector. Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. *Pharm. Res.* **9**:588–592 (1992).
3. J. W. A. Findlay, W. C. Smith, J. W. Lee, G. D. Nordblom, I. Das, B. S. DeSilva, M. N. Khan, and R. R. Bowsher. Validation of Immunoassays for bioanalysis: A pharmaceutical industry perspective. *J. Pharm. Biomed. Anal.* **21**:1249–1273 (2000).
4. C. M. Riley and T. W. Rosanke. Development of validation of analytical methods: progress in pharmaceutical and biomedical analysis (vol 3) Elsevier (Pergamon), NY 1996.
5. V. P. Shah, K. K. Midha, S. Dighe, I. J. McGilveray, J. P. Skelly, A. Yacobi, T. Layloff, C. T. Viswanathan, C. E. Cook, and R. D. McDowall. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetics studies. Conference Report. *Eur J Drug Metabol Pharmacokinetics* **16**:249–255 (1991).
6. Guideline on validation of analytical procedures: definitions and terminology International Conference of Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use Geneva 1995 (1996).
7. V. P. Shah, K. K. Midha, J.W.A Findlay, H. M Hill, J. D. Hulse, I. J McGilvary, G. McKay, K. J. Miller, R. N. Patnaik, M.L. Powell, A. Tonnelli, C. T. Viswanathan, and A. Yacobi. Bioanalytical method validation. A revisit with a decade of progress. *Pharm. Res.* **17**:1551–1557 (2000).
8. K. J. Miller, R. R. Bowsher, A. Celniker, J. Gibbons, S. Gupta, J. W. Lee, J. S. J. Swanson, W. C. Smith, and R. S. Weiner. Workshop on Bioanalytical Methods Validation for Macromolecules: Summary Report. *Pharm. Res.* **18**:1373–1383 (2001).
9. Guidance for the Industry. Bioanalytical Method Validation US Department of Health and Human Services FDA (CDER) and (CVM) May 2001.
10. J. O. Westgard. Points of care in using statistics in method comparison studies. *Clin. Chem.* **44**:2240–2242 (1998).
11. H. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M. C. Laparra, M. Laurentie, and J. C. Nivet. The SFSTP guide on the validation of chromatographic methods for drug analysis: from the Washington Conference to the laboratory. *Analytica Chimica Acta.* **391**:135–148 (1999).
12. R. Kringle and D. Hoffman. Stability methods for assessing stability of compounds in whole blood for clinical bioanalysis. *Drug Info J.* **35**:1261–1270 (2001).
13. U. Timm, M. Wall, and D. Dell. A new approach for dealing with the stability of drugs in biological fluids. *J. Pharm. Sci.* **74**:972–977 (1985).
14. D. Rodbard, Y. Feldman, M. L. Jaffe, and L. E. M. Miles. Kinetics of Two-Site Immunoradiometric (Sandwich) Assays-II. *Immunochem.* **15**:77–82 (1978).
15. B. D. Plikaytis, P. F. Holder, L. B. Pais, S. E. Maslanka, L. L. Gheesling, and G. M. Carlone. Determination of parallelism and nonparallelism in bioassay dilution curves. *J. Clin. Microbiol.* **32**: 2441–2447 (1994).
16. R. L. Placket and J. P. Burman. The design of optimum multifactorial experiments. *Biometrica* **33**:305–325 (1946).
17. J. M. Bland and D. G. Altman. Measuring agreement in method comparison studies. *Stat Meth Med Res* **8**:135–160 (1999).
18. C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Venkeerberghen, and D. L. Massart. Reappraisal of hypothesis testing for method validation; Detection of systematic error by comparing the means of two methods or two laboratories. *Analytical Chem.* **67**:4491–4499 (1995).
19. S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components* Chapter 3. John Wiley & Sons, Inc, New York, NY (1992).
20. R. W. Mee. β-expectation and β-content tolerance limits for balanced one-way ANOVA random model. *Technometrics* **26**:251–254 (1984).