

Workshop Report

Bioanalytical Method Validation for Macromolecules in Support of Pharmacokinetic Studies

JoMarie Smolec,^{5,21,22} Binodh DeSilva,¹ Wendell Smith,² Russell Weiner,³ Marian Kelly,⁴ Ben Lee,⁶ Masood Khan,⁷ Richard Tacey,⁸ Howard Hill,⁹ and Abbie Celniker¹⁰

CONTRIBUTORS: Vinod Shah,¹¹ Ronald Bowsher,¹² Anthony Mire-Sluis,¹ John W. A. Findlay,¹³ Mary Saltarelli,¹⁴ Valerie Quarumby,¹⁵ David Lansky,¹⁶ Robert Dillard,¹⁷ Martin Ullmann,¹⁸ Stephen Keller,¹⁹ and H. Thomas Karnes²⁰

Received January 20, 2005; accepted May 10, 2005

The development and validation of ligand binding assays used in the support of pharmacokinetic studies has been the focus of various workshops and publications in recent years, all in an effort to establish a guidance document for standardization of these bioanalytical methods. This summary report of the workshop from 2003 focuses on the issues discussed in presentations and notes points of discussion and areas of consensus among the participants.

KEY WORDS: bioanalytical assay; biological matrices; immunoassay; method validation.

INTRODUCTION AND HISTORICAL PERSPECTIVE

In recent years, interest in macromolecular therapeutics and development and evolution of analytical technologies have generated a need for standardization of bioanalytical methods validation procedures for macromolecules. This issue was addressed in a workshop on macromolecule methods validation held in 2000 (1); however, no guidance document was issued from this workshop. It was clear that

more in-depth discussions were needed to address the unique aspects of ligand-binding assays and that additional opinions needed to be considered to reach harmonization on bioanalytical methods validation. In response to this need, the American Association of Pharmaceutical Scientists sponsored a workshop entitled "Bioanalytical Method Validation for Macromolecules in Support of Pharmacokinetic Studies," which was held in Washington, DC, on May 12–13, 2003. The focus of this workshop was on quantitative ligand-binding assays, and its goal was to provide a framework for the development of a guidance document, which would be based on the publication that appeared in November 2003 (2) and the outcome of this workshop. The publication represented a committee's best effort at capturing industry recommendations for the validation of quantitative ligand-binding assays. The objectives of the 2003 workshop were as follows:

1. Address the need for "how to" information to validate methods used in pharmacokinetic studies and to generate the necessary documentation for methods development and validation;
2. Address validation issues relating to full, partial, and cross-validation;
3. Address run set up, with regard to standards and quality control samples (QCs), in order to:
 - (a) Reach agreement on running replicates to improve accuracy
 - (b) Reach agreement that batches must have a standard curve and QCs;
4. Define scientifically meaningful and valid acceptance criteria;
5. Move towards achieving global harmonization of methods validation.

¹ Amgen Inc., Thousand Oaks, California, USA.

² Bowsher Brunelle Smith, Greenfield, Indiana, USA.

³ Bristol-Myers Squibb, Princeton, New Jersey, USA.

⁴ Centocor, Malvern, Pennsylvania, USA.

⁵ Alta Analytical Laboratory, San Diego, California, USA.

⁶ Pfizer Global Research and Development, Ann Arbor, Michigan, USA.

⁷ Covance Laboratories, Chantilly, Virginia, USA.

⁸ PPD Development, Richmond, Virginia, USA.

⁹ HLS, Alconbury, Cambridgeshire, UK.

¹⁰ Millennium Pharmaceuticals, Cambridge, Massachusetts, USA.

¹¹ Food and Drug Administration, Rockville, Maryland, USA.

¹² LINCO Diagnostic Services, St. Charles, Missouri, USA.

¹³ Pfizer Global Research, Groton, Connecticut, USA.

¹⁴ Abbott Laboratories, Abbott Park, Illinois, USA.

¹⁵ Genentech, San Francisco, California, USA.

¹⁶ Lansky Consulting LLC, Burlington, Vermont, USA.

¹⁷ Takeda Pharmaceuticals N.A., Lincolnshire, Illinois, USA.

¹⁸ MDS Pharma Services, Zurich, Switzerland.

¹⁹ Protein Design Labs, Inc., Fremont, California, USA.

²⁰ Virginia Commonwealth University, Richmond, Virginia, USA.

²¹ 3985 Sorrento Valley Blvd, Ste C, San Diego, California 92121, USA.

²² To whom correspondence should be addressed. (josmolec@alta-immunochem.com.)

The proceedings of the 2003 workshop are summarized in this paper, which describes the material and information covered by the speakers at the workshop as well as discussion points or areas of consensus among the workshop speakers and audience. As such, this paper provides background on the committee's deliberations that resulted in the publication. This workshop summary is divided into sections representing areas and topics presented at the workshop, with supporting material added from the published document (2).

ISSUES IN DRAFTING RECOMMENDATIONS FOR VALIDATION OF IMMUNOASSAYS FOR BIOANALYSIS OF MACROMOLECULES

Various issues arise and need to be addressed at different phases of the development and validation of immunoassays for macromolecules. An assay life cycle can be categorized into three phases, each with unique issues and objectives: method development, prestudy validation, and in-study validation. The workshop publication was designed to address the various validation issues as they relate to each of the three phases of the assay life cycle.

There was general agreement that information related to the following assessments should be generated during method development:

- critical assay reagent selection and stability;
- assay format selection;
- diluents, plates, detection/system;
- standard curve model selection;
- matrix selection;
- specificity of the reagents;
- sample preparation;
- preliminary stability assessment; and
- preliminary assessment of assay robustness.

A validation plan developed during the prestudy validation phase should include a description of the intended use of the method (studies for which the assays will be used, length and size of the studies) and a summary of the performance parameters to be validated (standard curve, precision and accuracy, range of quantification, specificity and selectivity, stability, dilutional linearity, robustness, batch size, and run acceptance criteria), followed by a comprehensive report.

During the in-study validation phase, cumulative standard curve and QC data tables containing appropriate statistical parameters should be generated and included, along with the study sample values, in the final study report.

Prior to the publication of the committee's recommendations, published guidance documents on assay validation were limited to small molecules. There are key differences between conventional chromatographic methods and ligand-binding assays. Nonetheless, workshop participants agreed that, wherever possible, the recommendations for the validation of macromolecule assays should reflect the guiding principles of the small molecule assay guidance documents and attempt to achieve the greatest possible harmonization of validation methods.

Other issues addressed during the workshop include consensus definitions of key terminology (e.g., validation samples, QCs, run acceptance, method acceptance, specific-

ity, selectivity, cross-reactivity, robustness and ruggedness, sensitivity) and definition of the processes comprising a validation. These processes include characterization of reference standards, assay format, specificity and selectivity, standard curve acceptance, precision and accuracy, range of quantification, dilutional linearity, assessing parallelism, stability assessments, robustness and ruggedness, interference, as well as the rejection and acceptance criteria around each parameter.

The workshop participants noted the following practice discrepancies that required debate and harmonization:

- Defining the assay range using the validation samples vs. standard curve
- Determining the number of replicates of QCs and relating this number to the number of sample replicates
- Defining a batch/run as a set of standards and QCs
- Conducting parallelism "in study"
- Total error and the use of confidence intervals
- Standard curve editing rules
- Partial vs. full validation in matrix-substitution situations

The publication was noted to have two audiences, new associates and experienced scientists/analysts, who are interested in the detail behind the statistics. Such a diverse group makes it necessary to offer procedural recommendations. The goal of these recommendations is to help scientists develop ligand-binding assays that can be validated.

There was some agreement as to the information that should be included in a "how to" paper regarding immunoassay development and validation. A certain "closeness" or similarity to the small molecule guidance was considered desirable with respect to specifics on QC acceptance criteria for sample analysis and criteria for setting the range of quantitation for sample analysis. Thus, the goal of the workshop attendees was to produce a separate guidance document that is similar to the small molecule guidance, but contains more specific information related to immunoassays for macromolecules.

The lack of software tools that deal with nonlinear calibration and standard curve editing was discussed. Appropriate software can sometimes substitute for the lack of understanding and training in statistics or supplement a limited understanding; however, choices for software are limited. The guidelines developed from the workshop needed to include information to help users choose the best software for their needs and abilities.

REGULATORY PERSPECTIVES

The expectations of any regulatory agency with regard to validation of methods used to study biotechnology products are generally focused on two areas: following existing guidance documents and using sound scientific principles. However, for the validation of immunoassay-based pharmacokinetic assays of biologicals, there were no specific guidance documents available at the time of the workshop. Therefore, in preparing a new guidance document, it was important to understand the basic concepts of previously published guidances and to apply rational, scientifically guided approaches to their implementation.

The International Conference on Harmonization guidelines ICH Q2A and ICH Q2B contain general methods validation information (3,4), and the U.S. Food and Drug Administration guidance on bioanalytical methods validation (5) provides information that can be applied to a greater or lesser extent to ligand-binding assay validation, depending on the specific method and analyte.

Methods validation should take into account the purpose and requirements for the test, knowledge already gained about the product, and any associated risks. Methods utilized for good laboratory practices (GLP) studies, in-process control, and lot release testing for market approval should be validated, as should tests for determining clinical trial parameters.

Because pharmacokinetic assays are intended to provide data helpful in understanding the ADME (absorption, distribution, metabolism, and excretion) characteristics of the product, to help select the dose for preclinical and clinical studies, and to calculate the exposure of animals during toxicology studies, factors that could influence the validity of the data generated by the assay must be taken into account during method validation. There is a need for a clear understanding of the ability of the assay to detect the drug in the sample in which it is taken and the ways in which variables such as serum factors, endogenous material, soluble receptors, etc., affect the assay response. For these reasons, the validation of specificity/recovery, interference, etc. require a regulatory focus.

A thorough understanding of the biology of the product is needed to make certain that factors such as soluble receptors or other interfering substances are identified and studied in the assay. Of particular relevance is the effect of antibodies raised through immunogenicity on assay performance. Antibodies have been shown to affect detection in pharmacokinetic assays and, conversely, the drug may affect the detection of antibodies in immunogenicity assays. These are some of the factors that make quantitative assessments of macromolecules challenging.

There is also the very important question of whether the product detected in the assay is biologically active or not, because this affects understanding of exposure to fully functional material. It is necessary to correlate the response of the pharmacokinetic assay to that of a bioassay to determine how much biologically active material is being detected in the pharmacokinetic assay. The pharmacokinetic assay, in conjunction with bioassay studies, should also be examined for its ability to detect degraded product (monomers of multimeric proteins, clipped forms, etc.).

Levels of precision and accuracy do not necessarily have to be linked to any particular threshold, but should be scientifically justified so that the variability and accuracy of the assay are appropriate for its intended use (i.e., preclinical or clinical studies, exposure studies, dose-ranging studies, or comparability purposes). For example, because the ability to ensure accurate dosing is vital for a product with known toxicity or a narrow therapeutic window, this type of product may require an assay with higher precision and accuracy than a drug with a wider therapeutic window. This example points out the importance of ensuring that the performance characteristics of an immunoassay are suitable for its intended purpose, which can be determined through discussions with the end user—usually a pharmacokinetic analyst, clinician, or toxicologist.

Because of the complex nature of biological products, there is no single algorithm that can be used to address pharmacokinetic methods validation for each component of validation, but many of the existing parameters described in previously published guidance documents can be applied. In addition to the information from these guidance documents, immunoassay validation requires sound scientific justification of the steps taken during the validation process if it is to be deemed appropriate.

Differentiating the stages of validation, or levels of validation, is an empirical process, heavily dependent on the assay and its intended uses. It is often unclear where assay development ends, and qualification and validation begin. The term "qualification" appears in regulatory guidance documents in the context of the assessment of the suitability of equipment, in certain process validation test methods, and in comparability protocols for characterization tests. However, because there is no regulatory definition of test method qualification, adoption of such a term would require the rewriting of several guidance documents and a consensus as to the requirements for a method to be deemed "qualified" as opposed to validated.

There seems to be industry consensus that validation requires the writing of a validation protocol with predefined acceptance criteria. However, during the process of development/qualification, many of the factors necessary for validation, such as specificity, will have been addressed before making a decision to proceed with validation. It seems that a lot of repetition occurs under a validation protocol, including the testing of factors that should not change once the assay protocol is in place (e.g., cross-reactivity). Therefore, it may be worth considering the advantage of increasing quality assurance earlier during assay development, so that such standard information can be included in the final validation report without having to be repeated under a validation protocol.

In addition, the term "in-study validation" has appeared in certain documents. However, it may be worth considering that an assay should be validated before any study is started, with a standard operating procedure (SOP) and quality controls in place. Thus, the term in-study validation should be changed to "in-study monitoring" or "in-study confirmation" to make sure the predefined quality control parameters are being met.

BACKGROUND STATISTICS

The following procedural definitions were established and discussed:

- Intrabatch (within-run) precision is estimated by the pooled intrabatch standard deviation of measured concentration values from the calculated run means.
- Total random error or interbatch (between run) precision can be estimated by the standard deviation of all measured concentration values from the cumulative mean of all batches.
- Method accuracy, expressed as %RE (relative error, % bias) is determined by the percent deviation of the weighted sample mean from the sample nominal reference value.

- The weighted mean and sample overall mean are equal when the number of replicates is the same for all batches.

MACROMOLECULE REFERENCE STANDARD

The reference material to be used in an immunoassay needs to be carefully documented with regard to its source and characteristics. When possible, standard preparation, validation, and QC sample preparation should be performed using separate aliquots of single-use vials of the same source material. Lot numbers, purity, batch numbers, storage, stability, handling, and supporting documentation should be carefully monitored.

NONLINEAR CALIBRATION

Method Development Phase

A minimum of ten nonzero standard points in duplicate is recommended for the early characterization of a concentration–response relationship fit using the 4/5 parameter logistic (PL) function. Weighting 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. The recommendation in the publication is that a minimum of three independent runs should be analyzed to establish a calibration model, with appropriateness of the model to be judged by analysis of the %RE for back-calculated standard points ($\leq 20\%$). For a model to be considered acceptable, accumulated back-calculated values from all curves should have an absolute mean relative error of $\leq 10\%$ and a precision of $\leq 15\%$ for all concentrations in the range.

Prestudy Validation Phase

During prestudy validation, the recommendation is for a minimum of six nonzero standards, spaced evenly on a log scale, in duplicate within the anticipated range for a 4/5 PL function. The regression model should be confirmed in a minimum of six independent runs, the same runs in which method precision and accuracy are assessed. For a curve to be acceptable, the %RE of the back-calculated value for at least 75% of the standard points, not including anchor points, should be within 20% of nominal [25% at the lower limit of quantitation (LLOQ)]. The cumulative RE and CV should be $\leq 15\%$ for each standard ($\leq 20\%$ at LLOQ).

In-study Validation Phase

It is recommended that the standard curve be monitored during in-study validation. A standard point may be edited from a curve using the same criteria established during prestudy validation, with editing independent of and completed before assessment of QC performance. The final number of standards remaining after editing must be either $\geq 75\%$ of the total or a minimum of six standards in addition to the anchor points.

Calibration Curves and Models

The calibration curves for chromatographic assays usually demonstrate a direct and linear relationship between response and concentration of analyte. In contrast, immunoassays are inherently nonlinear.

The selection of the optimum calibration model is important in defining the correct quantification range, in maximizing accuracy and precision, and in enhancing the ability to achieve preset validation criteria and in-study quality control criteria. For an immunoassay calibration curve, the response–error relationship, defined as the variance in measurements of replicate responses, is a nonconstant function of the mean response (heteroskedasticity). The following points are important in choosing and maintaining a standard curve model:

- It is important for the concentration–response relationship for study samples and standards to be the same, with equal slopes and asymptotes.
- The dilution curves for the standard (the calibration curve should be prepared in the same matrix as the study samples) and study samples should also be parallel, and the zero concentration and maximum concentration responses should be the same.

Failure to achieve fundamental validity, which may compromise the accuracy and precision of the assay, is frequently caused by either “specific nonspecificity” (cross-reactivity) (interference caused by molecules structurally related to the analyte of interest) or “nonspecific nonspecificity” (matrix effect) (interference caused by matrix components not structurally related to the analyte).

The mathematical model most widely used to fit immunoassay calibration source data is the 4 PL model. If the calibration curve is asymmetric, inclusion of a fifth parameter may improve the data fit (5 PL). The mean, rather than individual responses, should be fitted to the validation model. Other calibration algorithms may be used (logit-log, cubic spline, etc.) if they demonstrate goodness of fit, but it must be remembered that these models represent attempts to linearize inherently nonlinear relationships. The selection of the validation model should be made during assay development and before validation experiments begin.

Proper weighting of the points in the calibration curve is also important to minimize bias and imprecision of interpolated values near the LLOQ and upper limit of quantitation (ULOQ) values. Replicates with smaller variances (normally in the pseudolinear portion of the curve) are given greater weight than those with larger variances (normally at asymptotes).

Goodness of fit should be assessed in at least three runs, with no more than 10% difference observed between the actual and back-calculated concentration values within the expected validation concentration range.

Recommendations for designing sigmoidal calibration curves for ligand-binding assays are as follows:

- Use at least six standard concentrations run at least in duplicate;
- Standards should be approximately equally spaced on a log scale;

- Use anchoring points beyond the LLOQ and ULOQ to assess the improvement of the overall fit;
- Evaluate positional effects.

Curve Editing

Consensus recommendations concerning the curve editing rules for prestudy vs. in-study validation are that curve editing rules need to be established *a priori*; all prestudy runs should be accepted and only rejected based on standard curve failure; and any curve editing during in-study validation should be based on target acceptance criteria for the fit of the standard points. If there are variability issues due to the asymptote, there should be anchor points at both ends.

PRECISION AND ACCURACY

In general, the workshop participants tended to agree with previously published recommendations for determining precision and accuracy.

Spiked QC samples are used throughout the life of an assay, from early assay assessment to define the assay range and to control the variability of the assay over time during in-study monitoring.

Method Development and Prestudy Validation Phases

During assay development, spiked QCs should be evaluated in a minimum of three development runs, with concentrations spanning the range of the standards with at least duplicate determinations for each concentration in each run. Target limits should be set at 20% (25% at the LLOQ) for cumulative %CV and absolute mean RE for each concentration.

During method development and prestudy validation, QC samples were termed validation sample/specimens (VSs). Recommendations involving VSs are as follows:

- VSs are prepared at five or more concentrations that span the range of the standard curve (LLOQ, $<3 \times$ LLOQ, midrange, and between the second and third uppermost standard curve points).
- During prestudy validation, there should be at least six runs (over several days) with VSs in duplicate. (At least two independent determinations per run.) A target limit of 20% (25% at LLOQ) for the cumulative %CV and %RE at each concentration is suggested.
- Interassay precision (%CV) and absolute mean bias (%RE) should both be $\leq 20\%$ (25% at LLOQ), with the sum of the %CV and absolute %RE to be less than 30%.
- VSs are used to define the LLOQ and ULOQ of the assay.

For in-study run acceptance, the recommendation is for at least two-thirds of all QCs to be within a specific percent of the corresponding nominal reference values, with at least 50% of the results within the specified limit for each QC sample. The publication also recommends

adoption of the 4–6–30 rule (see “In-Study Validation: Quality Control for Analysis of Test Samples”).

During the workshop, there was some debate as to whether or not the QC samples should be set at the LLOQ and ULOQ during the study validation.

Discussion Points

During the discussion that followed the workshop presentations for this session, the following items were addressed and agreements were reached:

- There was a consensus that the low and high QC samples need not be at the LLOQ and ULOQ, respectively, during sample analysis.
- There was agreement that a low QC specimen should be at $2 \times$ or $3 \times$ more than the LLOQ.
- There was agreement that VSs must be run over multiple days (six runs over multiple days) for good estimates of accuracy and precision in prestudy validation. More than three VSs should be used.
- No consensus was reached regarding use of the term “validation samples” or “validation specimens” (i.e., VSs) to refer to the prestudy validation phase and the term “QCs” to refer to in-study validation.
- The range of quantification is from the LLOQ to the ULOQ.
- There was no consensus reached as to whether the LLOQ and ULOQ have to be at the same concentrations as the limits of the standard concentrations.
- It was stated that the acceptance criteria should be the same for the prestudy and in-study validation phases.

SELECTIVITY AND SPECIFICITY

Discussions of selectivity and specificity generated the following consensus definitions for these terms:

- Specificity—the ability of an antibody to bind solely to the antigen of interest
- Selectivity—the ability of an analytical procedure to measure the analyte of interest in the presence of other sample constituents.

Specificity and selectivity evaluations are used to verify that a quantitative bioanalytical method is specific for the intended analyte, and that it can select the analyte from a complex biological matrix without positive or negative interference.

During method development, the specificity of the assay depends on the preestablished specificity of the antibody or antibody pairs. Data describing the binding characteristics of the antibody(ies) must be considered before selection. Assay specificity can be evaluated by spiking the sample matrix with variant forms of the analyte or with coadministered compounds. Evaluation of selectivity is performed by spiking multiple lots (at least ten sources) of sample matrix at or near the LLOQ (and possibly at higher concentrations) and assessing %RE. The recommended target acceptance criterion for selectivity during prestudy validation is that at least 80% of the matrices evaluated need to be within an

acceptable recovery of plus or minus 20–25% of the nominal or expected concentration.

Selectivity and specificity experiments should be repeated during in-study validation when relevant disease state matrices become available and if there is substitution of a lot of antibody.

DILUTIONAL LINEARITY/PARALLELISM

Because the quantitation range for an immunoassay may be very narrow, it is necessary to show that if the macromolecular analyte is present in concentrations above the range of quantification, it can first be diluted then accurately measured by the assay. Thus, dilutional linearity experiments need to be conducted. These experiments will also allow the detection of a possible prozone or “hook” effect.

Dilutional linearity can be assessed with spiked QC samples (at 100- to 1,000-fold higher concentrations than the ULOQ) that are diluted into an assay matrix obtained from an individual or a pool of individuals. During prestudy validation, dilutions should be confirmed. The back-calculated concentration of each diluted sample should be within 20% of the nominal or expected value. The precision of the cumulative back-calculated concentration should be $\leq 20\%$. During in-study validation, if a study sample needs to be diluted at a concentration higher than that assessed during prestudy validation, dilutional linearity should be repeated or a dilutional QC sample can be included in the assay.

Parallelism is assessed with multiple dilutions of actual study samples or with a sample representing the same matrix and analyte combination that will be generated during a study. It is recommended that the %CV between samples in a dilution series be $\leq 30\%$.

Stability Assessments

Stability experiments must mimic, as best possible, the conditions under which study samples will be collected, stored, and processed. Formal stability evaluations must be conducted with an established assay during prestudy validation. Stability samples must be prepared by spiking the analyte of interest, at high and low concentrations, into the same matrix as the study samples.

It is recommended that assessments be performed for bench-top stability, refrigerator temperature stability, whole-blood stability, freeze–thaw stability (three cycles, with no less than 12 h between thaws), and long-term freezer stability.

A standard curve and QC samples that are within expiration or freshly prepared should be used as the reference for comparison with the stability samples, which should employ the same acceptance criteria as QC samples. Alternative assessments may be applied (such as confidence intervals). Stability assessments continue during in study validation, with any changes in conditions or storage of samples verified before sample analysis.

The workshop encouraged the inclusion of intermediate time points for stability experiments, rather than a single sample at the start ($t = 0$) and the last day of the stability period. Also, the use of multiple replicates at each time point and replicate time points in different analytical

runs were recommended to generate more reliable data. The data obtained can then be represented by the average across the runs at each time point. It is thus unnecessary to rely on or use criteria based on individual results (e.g., the 4–6–20 rule).

Model- or semimodel-based approaches were suggested to greatly improve the characteristics of the stability study and to allow the analyst to better refute aberrant trends without inducing the bias associated with reanalysis. The use of confidence intervals was also encouraged.

CROSS-VALIDATION

A full validation involves method development, prestudy validation, and in-study validation.

A partial validation is conducted in situations in which method changes are considered to be minor in nature. Depending on the situation, various changes—including method transfer (might require comparison of results on blinded samples), changes to anticoagulant, changes in method (e.g., change in critical reagent), sample processing changes, changes in sample volumes, extension of the concentration range, selectivity issues, conversion from manual to automated method—may require validation experiments ranging from a single run to nearly a full validation.

A cross-validation is conducted when two validated methods are used within the same study or submission. The publication recommends that test samples be used to cross-validate the methods and data be evaluated with appropriate predefined acceptance criteria or statistical methods.

The workshop raised the issue of validating an assay for two reference standards, which means two validations. For comparison studies, both standards should be run. For a substitution, it is necessary to reestablish the assay parameters for the new standard. Assessment of the relative recovery of the previously analyzed samples will help readers to understand the relationship between the two reference standards.

ROBUSTNESS AND RUGGEDNESS

A “rugged” assay was defined as an assay that performs consistently during unavoidably differing operational conditions (e.g., changing analysts, laboratories, batch size, different instruments, days, or environmental factors). Therefore, assay ruggedness is determined by the assay’s consistency when the implementation of routine changes results in different operational conditions.

A “robust” assay is any assay that is able to withstand small, deliberate changes that may impact the assay (e.g., such as changes in incubation temperatures and/or times, light exposure, lot-to-lot difference of critical assay reagents, or changes to other elements of the SOP). Therefore, robustness of an assay is determined by the assay’s consistency when changes are implemented. All changes—which may include changes in incubation temperatures, light exposure, or matrix—must be tested and documented.

Assessing robustness and ruggedness during method development and validation provides reasonable assurance that assay performance will be acceptable over the range of

conditions an assay may encounter during use. During method development, variables and varying conditions need to be assessed for robustness and ruggedness to ensure that the subsequent validation is conducted within the limits set for these parameters. During prestudy validation, 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. Monitoring of QC performance and intra- and interassay precision at the end of studies will also provide information on the robustness and ruggedness of an assay conducted under different conditions.

The extent of the assessment of robustness and ruggedness depends on the anticipated application of the method, where the method is in its life cycle, available guidance and industry standards, and experience and common sense. The majority of robustness and ruggedness testing should be conducted during the method development phase to facilitate early identification of factors that may impact assay performance. Prestudy robustness and ruggedness validation should be limited to a few parameters demonstrating acceptable performance under predicted in-study conditions (e.g., incubation time tolerances, multiple analysts, varying batch sizes). Acceptable robustness and ruggedness are assumed once in-study validation data yield acceptable QC performance over the course of sample analysis.

IN-STUDY VALIDATION: QUALITY CONTROL FOR ANALYSIS OF TEST SAMPLES

Run acceptance is primarily based on the performance of the QC samples. At least four of six (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. This recommended 4-6-30 rule imposes limits simultaneously on the allowable random error (imprecision) and systematic error (mean bias). If an assay requires QC target acceptance using limits that differ from the 30% deviation from the nominal value, 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.

The recommendations for run acceptance criteria for macromolecules should be consistent with the guidance for small molecules whenever practical, except in cases where the small molecule guidance is scientifically incorrect. The current small molecule guidance is inconsistent in that the prestudy validation criterion of 15/20% would result in a large number of rejects during in-study validation. If nominal concentrations are the basis for comparison and the number of replicates is small, then a total error criterion should be used. If a well-established mean is the basis for comparison, then a criterion comparable to the precision limit should be used.

It should be noted that the limits of quantitation change with drift and may need to be reevaluated from time to time. High imprecision can be corrected and quantification limits

lowered by employment of replicate analysis. Replication will reduce an observed CV by a factor of $1/\sqrt{n}$.

The number of sample replicates should equal the number of QC sample replicates. The number of QCs in a batch should be a consistent percentage of the batch size to sustain the same power for detecting errors in large and small batches. The sequence of controls within a batch should be optimized so that the number of samples run between controls is minimized. The various QC sample concentrations should be evenly distributed throughout the run to provide optimal error detection at different concentrations. The error detection power of an "in-study" QC procedure can be increased either by increasing the number of control samples in a batch or by tightening the acceptance criteria.

The 4-6- X rule (where X is the selected percent deviation from nominal value) should also be used for other validation criteria, including selectivity, calibration residuals, dilution, and parallelism studies. The workshop recommended a confidence interval approach to QC monitoring as an alternative to the fixed 4-6- X rule. Finally, assuming the total error is X , and that X matches the prestudy criteria, the 4-6- X rule should be used for sample analysis QC acceptance.

There was consensus that the QCs should be used to demonstrate that the method is "in control." The high QC sample concentration should be in the upper quartile of the range, with the best practice putting it between the second and third uppermost standards (nonanchor points). The low QC sample concentration should be between the LLOQ and three times the LLOQ, with the best practice putting it between the second and third lowermost standards. If the low QC sample concentration is at the LLOQ and the low standard is lost, there is a risk of losing the run. The workshop also recommended that QC samples be scattered throughout the plate in a given run.

REFERENCES

1. 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).
2. B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tacey, H. Hill, and A. Celniker. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm. Res.* **20**(11):1885-1900 (2003).
3. International Conference on Harmonization. ICH Q2A. Text on Validation of Analytical Procedures. Federal Register. 1995;60 FR 11260. <http://www.fda.gov/cder/guidance/ichq2a.pdf>.
4. International Conference on Harmonization. ICH Q2B. Validation of Analytical Procedures Methodology. Federal Register. 1997;62 FR 27463. <http://www.fda.gov/cder/guidance/1320fnl.pdf>.
5. US Department of Health and Human Services. Draft Guidance for Industry: Analytical Procedures and Methods Validation, Chemistry, Manufacturing and Controls Documentation. Rockville, MD: US Dept of Health and Human Services, Food and Drug Administration. Aug 2000. <http://www.fda.gov/cder/guidance/2396dft.pdf>.