
Commentary

Quantitative Bioanalytical Methods Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays

C. T. Viswanathan,¹ Surendra Bansal,² Brian Booth,³ Anthony J. DeStefano,⁴ Mark J. Rose,^{5,11} Jeffrey Sailstad,⁶ Vinod P. Shah,⁷ Jerome P. Skelly,⁸ Patrick G. Swann,⁹ and Russell Weiner¹⁰

Received March 2, 2007; accepted March 8, 2007; published online April 26, 2007

Abstract. The Third AAPS/FDA Bioanalytical Workshop, entitled “Quantitative Bioanalytical Methods Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays” was held on May 1–3, 2006 in Arlington, VA. The format of this workshop consisted of presentations on bioanalytical topics, followed by discussion sessions where these topics could be debated, with the goal of reaching consensus, or identifying subjects where additional input or clarification was required. The discussion also addressed bioanalytical validation requirements of regulatory agencies, with the purpose of clarifying expectations for regulatory submissions. The proceedings from each day were reviewed and summarized in the evening sessions among the speakers and moderators of the day. The consensus summary was presented back to the workshop on the last day and was further debated. This communication represents the distillate of the workshop proceedings and provides the summary of consensus reached and also contains the validation topics where no consensus was reached.

KEY WORDS: bioanalytical; bioequivalence; LC-MS/MS; ligand binding; validation.

INTRODUCTION

Bioanalysis, employed for the quantitative determination of drugs and their metabolites in biological fluids, plays a significant role in the evaluation and interpretation of bioequivalence, pharmacokinetic (PK), and toxicokinetic studies. The quality of these studies, which are often used to support regulatory filings, is directly related to the quality of the underlying bioanalytical data. It is therefore

important that guiding principles for the validation of these analytical methods be established and disseminated to the pharmaceutical community.

The first American Association of Pharmaceutical Scientists (AAPS)/Food and Drug Administration (FDA) Bioanalytical Workshop in 1990 focused on key issues relevant to bioanalytical methodology and provided a platform for scientific discussions and deliberations. The workshop and the report (1), raised awareness of the need for validated bioanalytical methods for the regulatory acceptance of bioequivalence and pharmacokinetic data. Although the workshop addressed bioanalysis in general, it acknowledged the differences between chromatographic and ligand-binding (non-chromatographic based) methods. The workshop identified the essential parameters for bioanalytical method validation, i.e., accuracy, precision, selectivity, sensitivity, reproducibility, limit of detection and stability. The outcome of the first workshop and its report resulted in improved quality of data submissions to regulatory agencies.

Following the first workshop report (1) and the experience gained at the FDA, the draft Guidance on Bioanalytical Methods Validation was issued by the FDA in January 1999. This draft guidance provided stimulus and opportunity for further discussion at the 2nd AAPS/FDA Bioanalytical Workshop in January 2000. In addition, newer technology, such as chromatography coupled to tandem mass spectrometry (LC-

The views expressed in this article are those of the authors and do not reflect official policy at the FDA. No official endorsement by the FDA is intended or should be inferred.

¹ Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, Maryland 20857, USA.

² Hoffmann-La Roche Inc., Nutley, New Jersey 07110, USA.

³ Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, Maryland 20993, USA.

⁴ Procter & Gamble Pharmaceuticals, Mason, Ohio 45040, USA.

⁵ Amgen Inc., Thousand Oaks, California 91320, USA.

⁶ Sailstad and Associates, Durham, North Carolina 27707, USA.

⁷ Consultant, Gaithersburg, Maryland 20878, USA.

⁸ Consultant, Alexandria, Virginia 22309, USA.

⁹ Food and Drug Administration, Center for Drug Evaluation and Research, Bethesda, Maryland 20892, USA.

¹⁰ Bristol-Myers Squibb Inc., Pennington, New Jersey 08534, USA.

¹¹ To whom correspondence should be addressed. (e-mail: marose@amgen.com)

MS/MS), was discussed along with an update on ligand binding assays. This workshop resulted in a report "Bioanalytical Method Validation—A Revisit with a Decade of Progress" (2) and formed the basis for the FDA Guidance on Bioanalytical Methods Validation in May 2001 (3).

The evolution of divergent analytical technologies for conventional small molecules and macromolecules, and the growth in marketing interest in macromolecular therapies, has led to the workshop held in 2000 to specifically discuss bioanalytical methods validation for macromolecules. Due to the complexity of the issues, the workshop failed to achieve a consensus. To address the need for guiding principles for the validation of bioanalytical methods for macromolecules, the AAPS Ligand Binding Assay Bioanalytical Focus Group developed and published recommendations for the development and validation of ligand binding assays in 2003 (4).

As bioanalytical tools and techniques have continued to evolve and significant scientific and regulatory experience has been gained, the bioanalytical community has continued its critical review of the scope, applicability, and success of the presently employed bioanalytical guiding principles. The purpose of this Third AAPS/FDA Bioanalytical Workshop was to identify, review and evaluate the existing practices, white papers, articles and clarify the FDA Guidance. The workshop addressed quantitative bioanalytical methods validation and their use in sample analysis, focusing on both chromatographic and ligand binding assays.

GOALS AND OBJECTIVES

The purpose of the Third AAPS/FDA Bioanalytical Workshop was to:

- Review the scope and applicability of bioanalytical principles and procedures for the quantitative analysis of samples from bioequivalence, pharmacokinetic and comparability studies in both human and non-human subjects;
- Review current practices for scientific excellence and regulatory compliance, suggesting clarifications and improvements where needed;
- Review and evaluate validation and implementation requirements for Chromatographic and Ligand-based Quantitative Bioanalytical Assays, covering all types (sizes) of molecules;
- Review recent advances in technology, automation, regulatory and scientific requirements and data archiving on the performance and reporting of quantitative bioanalytical work; and
- Discuss current best approaches for the conduct of quantitative bioanalytical work regardless of the size of the molecule analyzed.

The Third AAPS/FDA Bioanalytical Workshop, held on May 1–3, 2006 in Arlington, VA, concluded with several recommendations to achieve the above goals and objectives. While the FDA guidance (3) remains valid, the recommendations obtained during the workshop were aimed at providing clarification and some recommendations to enhance the quality of bioanalytical work. This publication provides the clarification and recommendations obtained at the workshop with a view to achieve uniformity among the

practitioners and users of quantitative bioanalysis for all types of molecules.

NON-CHROMATOGRAPHIC ASSAY—SPECIFIC ISSUES

Differences between Ligand Binding Assays Supporting Macromolecule PK Analysis and Small Molecule Analysis by Chromatography

Ligand Binding Assays (LBAs) are used throughout many organizations attempting to discover or develop new chemical entities (NCE). Besides the obvious size difference between small and macromolecule analytes, there are key structural differences. Small molecules typically are organic molecules while macromolecules are complex biopolymers. In addition, small molecules are prepared by organic synthesis while macromolecules are typically formed biologically. As a direct result of how macromolecules are produced, the reference standards tend to be heterogeneous, often due to post-translational modification (e.g., glycosylation or phosphorylation). In contrast, small molecule reference standards are homogeneous with a high degree of purity. Generally, small molecules are often hydrophobic and macromolecules are often hydrophilic. While chemical stability is assessed for small molecules with relative ease, macromolecule stability assessment is generally more complex, requiring the evaluation of not only chemical and physical properties, but also biological integrity (i.e., is receptor binding affinity maintained?). Macromolecules are endogenous and/or structurally similar to endogenous counterparts, while small molecules are generally xenobiotics, foreign and not present in the sample matrix. The catabolism of small molecules is typically well defined while for macromolecules few specifics are known. Macromolecules typically have specific carrier proteins while small molecules can be generically bound to a number of endogenous proteins. Due to these significant differences between small and macromolecule analytes, different technologies, such as LC-MS for small molecules and LBAs for macromolecules, are often employed to determine drug levels for PK assessments.

Method validations for these divergent methods should consider important differences including the basis of measurement, the detection modality and whether a sample is measured directly in the matrix or extracted prior to analysis. The basis of measurement of LC-MS is owed to the chemical properties of the analyte, while for LBAs, the measurement depends on a high affinity biological binding interaction between the macromolecule analyte and another macromolecule(s) in the form of one or more capture/detection antibodies. Detection in LC-MS methods is direct and typically results in a linear measured response, where higher concentrations of analyte have a proportional increase in response. In contrast, the measured response in LBAs is indirect and this results in a non-linear, often sigmoidal, measured response. Owing to the characteristics of the assay system, the calibration standard curve range for a LC-MS method is broad, often covering several orders of magnitude. In contrast, the calibration range for an LBA is typically limited to less than two orders of magnitude. These analyte differences, combined with the unique technologies used to

measure analyte concentration, provide a strong rationale as to why consideration should be given to the need of employing some analyte specific (small versus macromolecule) method validation guidelines.

One major point of concern in discussing the method validations for these divergent technologies centers on standards and quality control (QC) acceptance criteria (i.e., the acceptable deviation from a nominal value expressed as a percentage). Current guidance recommends the 15/20 rule, where the first number, in this case 15%, is the acceptance criterion for all standards and quality control samples (QCs) with the exception of the lower limit of quantitation (LLOQ), where the acceptance criterion is increased to a 20% deviation. This rule was developed prior to the routine use of LC-MS, where chromatographic methods were employed, but internal standards were analyte analogues and not stable isotopes. When the 15/20 rule was proposed, most PK assessments that utilized LBAs (e.g., radioimmunoassay) measured small molecules. The typical radioimmunoassay (RIA) used high affinity polyclonal antibodies that were quite suitable to measure well characterized homogeneous organic small molecules. In most of these small molecule RIAs, meeting the 15/20 challenge was achievable and it is recommended that the 15/20 rule be continued when LBAs are used for small molecule analysis. However, nearly all small molecule analysis performed today is by LC-MS, often with the incorporation of a stable isotope internal standard; as a result, assay precision has continued to improve. In fact, the results of a method validation survey conducted for the Third AAPS/FDA Bioanalytical Workshop found that 89% of chromatography respondents used the 15/20 target.

As a result of small molecule analysis moving to the LC-MS platform, LBAs are now almost exclusively used to measure macromolecules. While some LBAs continue to be developed and validated to meet the 15/20 rule, different criteria are sometimes required because of the heterogeneous nature of macromolecules, and the fact that other macromolecules (antibodies) are employed in the assay. In fact, the Third AAPS/FDA Bioanalytical Workshop survey found that only 23% of the LBA respondents follow the 15/20 rule. Instead, 53% of respondents used somewhere between 20/25 (42%) and 30/30 (2%) as their acceptance rule, while 23% used "other criteria." These "other criteria" could possibly include statistically based approaches that estimate in-study assay performance based on pre-study validation results.

Ligand Binding Assays Pre-study Validation

During pre-study validation, method precision and accuracy are determined through the analysis of QCs (validation samples) prepared in a biological matrix equivalent to that anticipated for study samples. Due to the endogenous nature of some biopharmaceuticals, it may be necessary to deplete the matrix of the analyte or employ a 'surrogate' matrix to evaluate method accuracy and precision. One proposed validation protocol (4) recommends that matrix be spiked at 5 or more validation sample concentrations that span the range of quantification i.e., the anticipated lower limit of quantitation (LLOQ), ~3 times LLOQ, mid (geometric mean), high (~75% of the upper limit of quantitation

[ULOQ], and finally the anticipated ULOQ. As previously noted, the major sources of variability (imprecision and inaccuracy) differ based on technology. For LBAs, the inter-batch variance component is usually a greater contributor to the overall variability than the intra-batch variance component. It is recommended that at least two independent determinations be made for each validation sample per assay run across a minimum of six independent assays runs (balance validation design). For example, 12 reportable values would result from two measurements across six independent assay runs. An appropriate statistical method should then be used to compute the summary statistics (i.e., each validation sample, the repeated measurements from all runs should be analyzed together). A detailed description of this approach has been described previously (4).

For a method to be considered acceptable, it is recommended that both the inter-batch imprecision (%CV) and the accuracy, expressed as absolute mean bias (%RE) be $\pm 20\%$ (25% at LLOQ and ULOQ). As an additional constraint to control method error, it is recommended that the target total error [sum of the absolute value of the %RE (accuracy) and precision (% CV) be less than $\pm 30\%$ ($\pm 40\%$ at the LLOQ and ULOQ)]. The additional constraint of total error allows for consistency between the criteria for pre-study method validation and in-study batch acceptance. In assessing the acceptability of a method, including total error, it is not appropriate to reject assay runs. All assay runs during the validation should be included in the computation of summary statistics. The only exception would be runs rejected for cause or in cases where errors are obvious and documented.

Ligand Binding Assays In-study Acceptance Criteria

The recommended standard curve acceptance criteria for macromolecule LBAs are that at least 75% of the standard points should be within 20% of the nominal concentration (%RE of the back-calculated values), except at the LLOQ and ULOQ where the value should be within 25%. This requirement does not apply to 'anchor calibrators' which are typically outside the anticipated validation range of the assay and used to facilitate and improve 'sigmoidal' curve-fitting.

The recommended QC acceptance criteria for macromolecule LBAs includes the use of low, medium and high (LQC, MQC and HQC) QCs typically run in duplicate (i.e., 6 results=3 concentrations \times 2 reportable values per concentration), with assays being accepted based upon a 4-6-20 rule. Exceptions to this criterion should be justified (e.g., pre-study total error data approaching 30%). At least four of the six QCs must be within 20% of the nominal value. In addition, at least one QC sample per concentration needs to meet this criterion. If additional sets of QCs are used in a run, then 50% of them need to be 'in-range' at each concentration. The following are recommendations for the placement of the controls in relation to the standard curve range. The LQC should be placed above the second non-anchor standard, approximately 3 times the LLOQ. The MQC is placed near the mid point (geometric not arithmetic mean) of the standard curve, while the HQC should be placed below the second non-anchor point high standard and/or about 75% of the ULOQ.

CALIBRATION CURVE AND QC RANGES

QC samples serve to monitor the performance of the methodology throughout the course of the analysis. They are the basis for demonstrating, as required in 21 CFR 320.29(a), that the analytical method is sufficiently accurate, precise, and sensitive to measure the actual concentrations achieved in the body. For studies involving pharmacokinetic profiles spanning all or most of the calibration curve, three QC samples run in duplicate (or at least 5% of the unknown samples), spaced across the standard curve as per the FDA Guidance (3), are likely sufficient to adequately monitor method performance. For an analysis where the study data fall over a small percentage of the calibration curve, it is possible that none of the QC concentrations is near the concentrations of the unknowns, thus limiting the monitoring power of the QC samples.

If a narrow range of analysis values is known or anticipated prior to the start of sample analysis, it is recommended that either the standard curve be narrowed and new QC concentrations used as appropriate, or if the original curve is used, existing QC concentrations be revised or sufficient QC samples at additional concentration(s) added to adequately reflect the concentrations of the study samples. Narrowing of the standard curve and preparation of new QC samples requires only a partial validation to assure adequate performance of the new curve and QCs. A full validation is not required.

If a narrow range of analysis values is unanticipated, but observed after start of the sample analysis, it is recommended that the analysis be stopped and either the standard curve narrowed, existing QC concentrations revised, or QC samples at additional concentrations be added to the original curve prior to continuing with sample analysis. It is not necessary to reanalyze samples analyzed prior to optimizing the standard curve or QC concentrations.

CARRYOVER AND CONTAMINATION EVALUATION

Contamination, carryover, or blank response from matrix or reagents can affect the accuracy and precision of quantitation at all concentrations. However, low concentration samples are most affected as a percentage of concentration. Care should be taken to minimize interference from all contamination factors and the interference should not significantly affect the accuracy and precision of the assay.

Carryover does not necessarily involve only the next sample in the sequence. In fact, carryover from late eluting residues on columns may affect chromatograms several samples later. Carryover from residues in rotary sampling/switching valves often appears later in the samples. Precautions should be taken to avoid contamination during sample collection and preparation. Carryover should be assessed during validation by injecting one or more blank samples, after a high concentration sample or standard. The injector should be flushed with appropriate solvents to minimize carryover. If carryover is unavoidable for a highly retained compound, specific procedures should be provided in the method to handle known carryover. This could include injection of blanks after certain samples. Randomization of

samples should be avoided, since it may interfere with the assessment of carryover problems. Contamination can be assessed by monitoring blank response in the presence of high concentration samples or standards. The assay platform (manual or automated), configuration of sampling and extraction method (manual, automated, on-line or solid phase, etc.) in the assay should be taken into consideration when ascertaining contamination. There is no standard acceptable magnitude of carryover for a passing bioanalytical run. Carryover should be addressed in validation and minimized, and an objective determination should be made in the evaluation of analytical runs.

During validation, the operator should assess the analyte response due to blank matrix while eliminating or minimizing other contaminations. The analyte response at the LLOQ should be at least 5 times the response due to blank matrix. For immunoassays, and if the analyte is present endogenously in the matrix, the blank response can exceed 20% of LLOQ, but the contribution should not interfere with the required accuracy in the measurement of the LLOQ. In such cases, specific procedures should be provided in the method to handle blank matrix response.

DETERMINATION OF METABOLITES DURING DRUG DEVELOPMENT

A draft FDA Guidance for Industry, entitled "Safety Testing of Drug Metabolites" was issued in June 2005 by the Center for Drug Evaluation and Research (CDER) (5). There is general support from the pharmaceutical community for the idea that a more extensive characterization of the pharmacokinetics of unique and/or major human metabolites (UMMs) would provide greater insight into the connection between metabolites and toxicological observations. This information would be best generated by the use of rugged, bioanalytical methods applied at appropriate times in drug development.

Characterization of UMMs should proceed using a flexible, "tiered" approach to bioanalytical methods validation. This tiered approach would allow metabolite screening studies to be performed in early drug development using bioanalytical methods with limited validation, with validation criteria increasing as a product moves into clinical trials. A tiered validation approach to metabolite determination would defer bioanalytical resource allocation to later in the drug development timeline when there is a greater likelihood of drug success. As a minimum, the specifics of this tiered validation process should be driven by scientifically-appropriate criteria, established a priori.

INCURRED SAMPLE REANALYSIS

There are a number of situations where the performance of standards and QCs may not adequately mimic that of study samples from dosed subjects (incurred samples). Examples include, metabolites converting to the parent species, protein binding differences in patient samples, recovery issues, sample inhomogeneity, and mass spectrometric ionization matrix effects. These factors can affect both the reproducibility and accuracy of the concentration determined in incurred samples. While these effects are often

Table 1. Details of Documentation Desirable at the Analytical Site and in Validation and Analytical Reports

Items	Analytical Site	Validation Report	Validation Report Appendix	Analytical Report	Analytical Report Appendix
Standard	Certificate of analysis, purity, stability for analyte Record of receipt and storage Lack of interference between IS and analyte	Batch/Lot #, purity and manufacturer Stability at time of use		Batch/Lot #, purity and manufacturer Stability at time of use	
Stock solution preparation	Records of preparation Storage location & condition	Preparation dates Storage conditions		Storage conditions	
Calibrators and QCs preparation	Freezer log (sample ingress/egress, temperature)				
Run acceptance criteria	SOP for calibrators, QCs & chromatographic interferences	Short description	SOP (optional)	Short description	SOP (optional)
Assay procedure	SOP for the method	Brief description of method of extraction, and analysis	SOP (optional)	Brief description	SOP (optional)
Sample tracking	Study sample receipt, condition on receipt & location of storage Tracking of QC, calibrators and study samples Freezer logs	Storage condition & location		Dates of receipt of shipments and contents Sample condition on receipt Storage location & condition	
Analysis	Dates of extraction and analysis and instrument ID for each run Identity of QCs, calibrators & study samples Documentation of processing of calibrators, QCs and study samples for each run Documentation of instrument settings and maintenance Run summary sheets 100% chromatograms LIMS and Mode of integration Extraction dates	Table of runs, instrument ID & analysis dates Table of calibrator results of all runs with accuracy and precision Tables of within and between run QC results (accuracy and precision) Bench-top, freeze-thaw, long-term and post-preparative & stock solution stability data Extraction recovery & matrix effect	Representative chromatograms Cross-validation, if applicable Additional validation, if any Long-term stability appended or written in a separate report	Table of all runs, & analysis dates Table of calibrator results of all passed runs with mean & % CV Tables of QC results of all passed runs with accuracy and precision. O.K to include QC results of the failed runs.	Chromatograms from 5-20% of subjects for ANDA and Representative chromatograms for NDA submissions

Failed runs	Same as in "Analysis"	Identify runs, assay date and reason for failure	Identify runs, assay dates, and reason for failure
Reintegration	Audit trail: original & reintegration Reason for reintegration Mode of reintegration	Identify runs, assay date and reason for failure	SOP (optional)
Deviations from SOPs/method	Documentation of deviations and unexpected events Investigation of unexpected events Impact assessment	Description of deviations Impact on study results Description and supporting data of significant investigations	Description of deviations Impact on study results Description and supporting data of significant investigations
Reassay	Refer to "Analysis"	SOP (optional)	SOP Table of Sample IDs, reason for reassay, original and reassay values and Run IDs
Communication	SOP for reassay criteria Between analytical site and clinical site/sponsor		

SOP standard operating procedures, IS internal standard, QC quality control, QCs quality control samples, %CV interbatch imprecision, ANDA Abbreviated New Drug Application, NDA New Drug Application, LIMS Laboratory Information Management Systems

characterized and minimized during method development using QC samples, it is important to assure that they are under control when the method is applied to the analysis of incurred samples.

A proper evaluation of incurred sample reproducibility and accuracy needs to be performed on each species used for Good Laboratory Practice (GLP) toxicology experiments. It is not necessary for additional incurred sample investigations to be performed in toxicology species once the initial assessment has been performed. Incurred sample evaluations performed using samples from one study would be sufficient for all other studies using that same species.

It is generally accepted that the chance of incurred sample variability is greater in humans than in animals, so the following discussion pertains primarily to clinical studies. The final decision as to the extent and nature of the incurred sample testing is left to the analytical investigator, and should be based on an in-depth understanding of the method, the behavior of the drug, metabolites, and any concomitant medications in the matrices of interest. There should be some assessment of both reproducibility and accuracy of the reported concentration. Sufficient data should be generated to demonstrate that the current matrix produces results similar to those previously validated. It is recognized that accuracy of the result generated from incurred samples can be more difficult to assess. It requires evaluation of any additional factors besides reproducibility upon storage, which could perturb the reported concentration. These could include metabolites converted to parent during sample preparation or LC-MS/MS analysis, matrix effects from high concentrations of metabolites, or variable recovery between analyte and internal standard. If a lack of accuracy is not due to assay performance (i.e., analyte instability or interconversion) then the reason for the lack of accuracy should be investigated and its impact on the study assessed. The extent and nature of these experiments is dependent on the specific sample being addressed and should provide sufficient confidence that the concentration being reported is accurate.

The results of incurred sample reanalysis studies may be documented in the final bioanalytical or clinical report for the study, and/or as an addendum to the method validation report.

In selecting samples to be reassayed, it is encouraged that issues such as concentration, patient population and special populations (e.g., renally impaired) be considered, depending on what is known about the drug, its metabolism and its clearance. First-in-human, proof-of-concept in patients, special population and bioequivalence studies are examples of studies that should be considered for incurred-sample concentration verification. The study sample results obtained for establishing incurred sample reproducibility may be used for comparison purposes, and do not necessarily have to be used in calculating reported sample concentrations.

DOCUMENTATION ISSUES

Although the current guidance for the documentation section remains valid, further issues are now addressed and details are provided herein (Table I) to facilitate effective documentation. Records generated during the course of method validation and study sample analysis are source

records and should be retained to demonstrate the validity of the method under the conditions of use, and to support the statements made in the report. This is necessary to enable the reconstruction of the laboratory events as they occurred since the information generated by the individual laboratory might differ from what the sponsor includes in the application.

Documentation at the Analytical Site

1. Documentation of standard analyte can be done by Certificate of Analysis (CA) or recertification of purity or stability data at the time of the use. In case of the internal standard, no specific CA is necessary but the lack of interference between IS and analyte should be established.
2. The source data (run preparation, extraction and run summary sheets, and chromatograms) of all analytical and validation runs, including failed runs should be retained.
3. Reintegrated chromatograms should be explicitly identified. The reason for reintegration and the mode of reintegration should be documented. The original and reintegrated chromatograms should be retained ideally as electronic records.
4. Any problems during extraction and analysis (i.e., run interruption, clogging of columns) should be identified. The appropriate remedial action should be documented.
5. In the case of multi-analyte assays (simultaneously measurement of multiple analytes in each sample), when samples are reassayed only for one analyte (e.g., because the analyte failed to meet acceptance criteria in the original assay), the raw data collected for the other analytes should also be retained.

Analytical/Validation Reports Should Include:

- 1a. Summary Table of *all* analytical runs analyzed. The tables should list the runs with run IDs, dates of analysis, whether runs passed or failed, reason for the failure and any deviations from the validated method.
- 1b. Summary Table of *all* validation runs analyzed. The tables should list the runs with run IDs, dates of analysis, whether runs passed or failed, and the reason for the failure. QC data from validation runs that only failed to meet QC acceptance criteria with no assignable cause for failure should be included in the precision and accuracy estimation.
2. Deviations from SOPs and assay procedures, and significant unexpected events should be identified and their impact assessed.

Source Data Documentation

The actual conditions of use should be stated in documentation. For example, the source documentation for stability determinations during method validation should explicitly record experimental conditions such as storage temperature and duration, use of freshly prepared standard curves etc. Such documentation is necessary to confirm that

validation experiments support the storage conditions that existed during sample analysis.

Modification of calibration response (deletion of individual standard points that exceed predefined acceptance limits or alteration of the standard curve range) and QC levels (adding QCs or shifting in the concentration range of the study samples) should be documented with sufficient detail to demonstrate that the changes were justified and/or followed established procedures.

Regarding chromatographic methods, source documentation should include original and reintegrated chromatograms for accepted runs, along with the reason for changing integration parameters across a run or for individual samples within a run. Disabling electronic audit trails that record changes to integration parameters is not acceptable.

Final Report Documentation

A complete account of the performance of the bioanalytical method should be provided in the final report for both method validation experiments and study sample analysis. Although drug concentration data from the rejected runs need not be included in the final report, a brief description of the reasons and a tabular listing of rejected runs should be provided. The information provided would be helpful in the evaluation of the overall assay performance and acceptance of runs rejected and accepted. The final report should include a tabular listing of the actual QC results from all runs during method validation and accepted runs during study sample analysis. A table listing all reassayed samples, reason for reassay, and the values for original, reassay and final should be included in the final report.

Currently, as described in the FDA Guidance (3), 5–20% of all chromatograms, including QCs samples and standards, must be submitted with an NDA or ANDA. Because of the crucial nature of bioequivalence studies, the practice of submitting 20% of chromatograms from serially selected subjects should be continued for both NDAs and ANDAs. In general, representative chromatograms of typical analysis for other PK studies for NDAs should be sufficient for FDA submission. In circumstances where other PK studies are critical to the approval of the NDA, 20% of chromatograms may be requested for submission. However all original chromatograms and reintegrations should be retained at the site and available for audit if necessary.

Further reference is made to the FDA Guidance (3) that describes the final report attributes in detail and remains generally applicable.

STABILITY RECOMMENDATIONS

Drug stability experiments should mimic conditions under which samples are collected, stored and processed, as closely as possible. The experiments should be conducted in unaltered representative matrix, including the same type of anticoagulant. In cases where stripped or altered matrix is used for preparation of study calibration standards and/or QC samples, stability evaluation still must be conducted in samples prepared in unaltered matrix. If a stabilizer is normally employed with incurred samples, it should be employed with the stability samples also.

Short term stability experiments should be designed and conducted to cover the type of storage conditions that are to be expected for study samples. This generally includes an evaluation of three freeze/thaw cycles, 4-h bench top and refrigerated stability. During freeze/thaw stability evaluations, the freezing and thawing of stability samples must mimic the intended sample handling conditions to be used during sample analysis. If study samples are to be stored on wet ice, for thawed periods greater than 4 h, then these conditions should be evaluated during validation as well. If during the sample analysis for a study, a sample was thawed through greater than three cycles or if storage conditions changed and/or exceeded the sample storage conditions evaluated during method validation, stability must be established under these new conditions in order to demonstrate that the concentration values from these study samples are valid.

While short term stability measurements are generated during method validation, long term measurements are initiated during method validation, possibly evaluating analyte stability for a period of a few weeks, with the remaining long term storage time points evaluated post method validation. This post-validation data can then be added to the original validation data in the form of a validation report addendum or as a stand alone stability report. Long term stability should be evaluated at the expected storage conditions, including expected satellite storage temperature and duration (e.g., prior to shipment to the analytical laboratory). In consideration of this, there may be the need to include both -70 and -20°C evaluations (e.g., when samples are stored under different conditions at the various study locations). Refer to Section entitled “[Separate Stability Experiments Required at \$-70^{\circ}\text{C}\$ if Stability Shown at \$-20^{\circ}\text{C}\$](#) ” for additional discussion.

Stability evaluations should be performed against freshly prepared standard curves. When evaluating data generated from stability experiments, intended (nominal) concentrations should be used for comparison purposes. Additionally, in order to determine if the initial batch of stability samples are suitable for the subsequent stability experiments, a comparison with the initial day 0 or day 1 samples is recommended. If the measured concentration of the day 0 or 1 stability sample differs substantially from the intended concentration, this difference may be an indication that the bulk stability samples were not prepared correctly and preparing new bulk stability samples should be considered.

With respect to stock solutions prepared from certified reference standards, if the reference standard is within its expiration date when the stock solution is prepared, there is no need to prepare a new stock solution when the reference standard expires. When the stock solution exists in a different state (solutions vs. solid) or in a different buffer composition (generally the case for macromolecules), the stability data on this stock solution should be generated to justify the duration of stock solution storage stability. In general, newer stock solutions within their established stability period (e.g., a solution with established 60-day stability used on Day 55) should not be used to measure stability of an older solution (e.g., 120 days old). Although the newer stock may meet stability criteria for bioanalytical purposes, the chance of misinterpreting the stability of the older solution is high. The suggestion is to make a solution fresh from powder when determining the stability of any older stock.

An additional concern for LBAs is reagent stability. This includes, but is not limited to, antibodies, antibody conjugates (e.g., horseradish peroxidase, biotin and avidin conjugates). Therefore, during method validation, documentation should be made of the conditions under which the principal reagents maintain sufficient stability to meet the basic requirements of assay performance. Some of these data will need to be generated by the sponsor in the case of proprietary reagents, while other stability data can be obtained from the manufacturer for commercially available reagents. When using the manufacturer's data, reagents must be stored as recommended by the manufacturer. If different conditions are used, the analytical investigator will need to generate the appropriate storage stability data.

The evaluation of extract stability with a freshly prepared standard curve is not part of routine validation testing, but should be conducted as needed. In cases where extracted samples are stored prior to analysis (e.g., extracted samples are refrigerated for several hours or days prior to placement on the instrument), extract stability should be demonstrated for the storage temperature and duration. With regard to autosampler reinjection reproducibility, a freshly prepared standard curve is not necessary.

MATRIX EFFECTS FOR MS BASED ASSAYS

One phenomenon influencing mass spectrometry (MS) based bioanalytical assays is matrix effect. Matrix effect is the suppression or enhancement of ionization of analytes by the presence of matrix components in the biological samples. Quantitative measurement of matrix effect provides useful information in validation of MS based bioanalytical methods. The quantitative measure of matrix effect can be termed as Matrix Factor (MF) and defined as a ratio of the analyte peak response in the presence of matrix ions to the analyte peak response in the absence of matrix ions, i.e.,

$$\text{Matrix Factor} = \frac{\text{Peak response in presence of matrix ions}}{\text{Peak response in absence of matrix ions}} \quad (1)$$

A Matrix Factor of one signifies no matrix effects. A value of less than one suggests ionization suppression. An MF of greater than one may be due to ionization enhancement and can also be caused by analyte loss in the absence of matrix during analysis. Internal standard (IS) normalized MF is the MF of analyte divided by the MF for IS. The IS normalized MF can also be obtained by substituting peak response with peak response ratio (analyte/IS) in the above equation for MF. Stable isotope labeled IS minimizes the influence of matrix effects most effectively since the matrix effects observed for stable isotope labeled IS are generally similar to those observed for the matching analyte. Analog IS may also compensate for matrix effects; however the stable isotope labeled internal standards are most effective and should be used whenever possible.

An absolute MF (or IS normalized MF) of about one is not necessary for a reliable bioanalytical assay. However, highly variable MF in individual subjects would be a cause for the lack of reproducibility of analysis. To predict the variability of matrix effects in samples from individual

Table II. Routine Drug Analysis Process and Run Acceptance Criteria

Process or Criteria	Chromatographic Assays	Ligand Binding Assays
1. Preparation of standards and QC samples	Standards and QC samples can be prepared from the same spiking stock solution, provided the solution stability and accuracy have been verified. A single source of matrix may also be used, provided selectivity has been verified.	
2. Placement of samples	Standard curve samples, blanks, QCs, and study samples can be arranged as considered appropriate within the run, and support detection of assay drift over the run.	
3. Number of calibration standards in a run	<p>Include with each analytical batch: Blank matrix Zero standard (matrix sample with internal standard)</p> <p>Non-zero Calibration Standards: a minimum of six (6) standard points</p>	<p>Include with each analytical batch or micro-titer plate: Blank matrix Non-zero Calibration Standards: a minimum of six (6) standard points. Can include anchor points (below LLOQ or above ULOQ in the asymptotic low and high-concentration end of the standard curve)</p>
4. Acceptance criteria for calibration standards	<p>Residuals (absolute difference between the back calculated and nominal concentration) for each calibration standard should meet the following limits: LLOQ standard < 20% All other standards < 15%</p> <p>A minimum of 75% standards (at least 6 non-zero points) should be within the above limits for the analytical run to qualify. Values falling outside these limits can be discarded, provided they do not change the established model.</p>	<p>Residuals for each calibration standard should meet the following limits: LLOQ and ULOQ standards < 25% All other standards < 20% Any anchor points if used, are not to be included in the above acceptance criteria.</p>
5. Number of QC samples in a batch	<p>Include QC samples at the following three concentrations (within the calibration range) in duplicate with each analytical batch: Low: near the LLOQ (up to 3x LLOQ) Medium: midrange of calibration curve High: near the high end of range</p> <p>Each analytical batch should contain six or a minimum of 5% of the total number of unknown samples. Add QCs in multiples of three concentrations (low, medium, high) when needed.</p>	<p>QC samples at the following three concentrations (within the calibration range) in duplicate should be added to each micro-titer plate: Low: above the second non-anchor standard, approx. 3x LLOQ Medium: midrange of calibration curve High: below the second non-anchor point high standard at about 75% of ULOQ</p>
6. Acceptance criteria for QC samples ^a	<p>Allowed % deviation from nominal values: QCs prepared at all concentrations greater than LLOQ < 15% Low QC (if prepared at LLOQ) < 20%</p>	<p>Allowed % deviation from nominal values: QCs prepared at all concentrations other than LLOQ and ULOQ < 20% Low and high QC (if prepared at LLOQ or ULOQ) < 25% In certain situations wider acceptance criteria may be justified, e.g., when total error during assay validation approaches 30%</p>

<p>At least 67% (four out of six) of the QC samples should be within the above limits. 33% of the QC samples (not all replicates at the same concentration) can be outside the limits. If there are more than two QC samples at a concentration, then 50% of QC samples at each concentration should pass the above limits of deviation.</p>	<p>Accuracy can generally be improved by replicate analysis. Therefore, duplicate analysis is recommended. If replicate analysis is performed, the same procedure should be used for samples and standards.</p>
<p>7. Replicate analysis</p>	<p>In general samples can be analyzed with a single determination without replicate analysis if the assay method has acceptable variability as defined by the validation data. Duplicate or replicate analysis can be performed for a difficult procedure where high precision and accuracy may be difficult to obtain.</p>
<p>8. Multiple analytes in a run^b</p>	<p>Samples involving multiple analytes in a run should not be rejected based on the data from one analyte failing the acceptance criteria.</p>
<p>9. Rejected runs</p>	<p>The data from rejected runs need not be documented, but the fact that a run was rejected and the reason for failure should be reported.</p>

^a The issue of monitoring the effect of sample dilution is referred to in the FDA guidance (3). The guidance indicates that if the dilutions are conducted with like matrix (human plasma for human plasma), no within-study dilution matrix QC samples are necessary. However, the extent to which samples are allowed to be diluted should be tested during validation. If tested during validation, there is no need to run dilution QCs up to the tested dilution factor during sample analysis. If during sample analysis it is determined that the required dilution factor is greater than the extent tested during validation, an additional dilution factor should be tested during sample analysis. On the other hand, if the dilution is allowed and performed with an unlike matrix, QC samples should be diluted in the same manner as the study samples, and should be analyzed with the diluted samples. All diluted QCs should be created within the assay calibration range, and similar acceptance criteria as defined here should be used, unless alternate specific criteria can be justified.

^b Multiple analyte assessment in a single analytical run was a topic discussed at the meeting. Although the FDA Guidance (3) indicates that a run should not be rejected for the remaining analytes if one fails, it does not address how to assess and report all analyte concentrations upon re-analysis of the failed analyte(s). In this regard, concentrations from the first accepted run should be reported and if this analyte is repeated in simultaneous assays when analyzing for different analytes, it is not necessary to quantitate the already reported analytes. However, the source data from all acceptable runs, regardless of whether the concentrations from these runs were reported or not, should be retained.

QC quality control, *QC_s* quality control sample, *LLOQ* lower limit of quantitation, *ULOQ* upper limit of quantitation

subjects, determine the MF (or IS normalized MF) for six individual lots of the matrix. The variability in matrix factors, as measured by the coefficient of variation should be less than 15%. If the matrix is rare and hard to obtain, the requirement for assessing variability of matrix factors in six lots can be waived. Stable isotope labeled internal standards help by normalizing MF to a theoretical value of one, and thereby reduce the effective IS normalized MF variability. When using stable isotope IS, it is not necessary to determine the IS normalized MF in six different lots.

SYSTEM SUITABILITY

Scientifically qualified and properly maintained instruments should be used for implementation of bioanalytical methods in routine drug analysis. As part of qualifying instruments, performance of system suitability ensures that the system is operating properly at the time of analysis. System suitability checks are more appropriately used for chromatographic methods to ensure that the system is sufficiently sensitive, specific and reproducible for the current analytical run. However, the system suitability tests do not replace the required run acceptance criteria with calibration standards and QC samples. System suitability tests, when appropriate, are recommended to ensure success, but are not required, nor do they replace the usual run acceptance criteria.

REFERENCE STANDARDS

Analytical reference standards are used for the preparation of calibration standards and QC samples. Reference standard lot numbers, purity, storage, stability, handling, and supporting documentation should be monitored and maintained. Reference standards should be used prior to their expiration or re-certification dates. Some compounds used as internal standards or rare metabolites are available in very small amounts, and their certificates of analysis (CA) may not be available. If the full CA is not available for rare metabolites, at a minimum, the documented purity information should be obtained. CA or purity information of internal standards is not always necessary for the use of internal standards. When purity information is not available for the internal standard, it needs to be demonstrated that the internal standard does not interfere with the chromatography of the analyte(s) of interest. Macromolecular reference standards are often heterogeneous and may present unique comparability and stability considerations.

RUN ACCEPTANCE CRITERIA

The elements of chromatographic run acceptance previously described in the FDA Guidance (3) are listed in Table II. Recommendations specific to LBAs are also provided in “[Ligand Binding Assays In-study Acceptance Criteria](#)” section.

VALIDATION TOPICS WITH NO CONSENSUS

This section describes the topics for which no consensus could be reached during the Third AAPS/FDA Bioanalytical

Workshop. The purpose of this section is to provide comments and emphasize that further discussion and direction will be necessary.

Cross-validation of Bioanalytical Methods when Using Different Anticoagulant Counter-Ions

There was recognition of a distinct difference between EDTA and heparin containing plasma and that a bioanalytical method validated for one could not be used for the other without some revalidation of the method but no consensus was reached for the need for cross-validation when using the same anti-coagulant with a different counter-ion. As an example, attendees could not agree on the degree of cross-validation necessary for a method validated using sodium heparinized plasma when it was applied to a lithium heparinized sample.

Cross-validation Required when Using Different Strains or Sexes of a Species

No decisive arguments came forward in support of this activity although it was agreed that there could be some differences in biological matrix originating from the different strains or sexes. The general trend of the debate was that validation experiments to address such differences should not be considered the norm and should be performed when there are method related concerns that can be attributed to a specific strain or sex-related difference.

Cross-Validation Required when Moving a Method Between LC-MS/MS Instruments

Moving from different models of instrument (e.g., Sciex API-4000 to a Sciex API-5000) would require cross-validation, but there was very little support for requiring cross-validation when switching between equivalent instruments. It was noted that a qualification experiment is usually performed by most analysts before samples are run on a given instrument and that these experiments were usually sufficient to allow the qualification of a new instrument for a specific assay.

Specific Criteria for Cross-Validation

The term cross-validation was used liberally throughout the Third AAPS/FDA Bioanalytical Workshop when considering all different types of changes in bioanalytical methods. Cross-validation was discussed when using matrix from different species, when using matrix with different anticoagulants, when transferring methods to other laboratories and when transferring methods to other analysts. However no specific strategies, procedures and acceptance criteria were discussed to adequately perform these cross-validation experiments. Cross-validation procedures and acceptance criteria need to remain flexible, considering the various bioanalytical situations where it would be required. Specific cross-validation criteria should be established a priori via a standard operating procedure.

Separate Stability Experiments Required at -70°C if Stability Shown at -20°C

A significant group of workshop attendees believed that biological sample stability shown at a given temperature (e.g., about -20°C), automatically defined sample stability at a lower temperature (e.g., about -70°C) based on Arrhenius principles of chemical reactivity. These principles dictate lower rates of chemical reactivity (i.e., analyte degradation) occurring at lower temperatures. The argument against assuming sample stability at lower temperatures was based on matrix degradation rather than chemical stability of the analyte. An effective (if anecdotal) argument was made during the meeting which questioned the stability of biological matrix proteins at lower temperatures. It was argued that a lower temperature could cause denaturation or precipitation of matrix proteins and that this could affect protein binding or the ability to extract the drug from the matrix. There was general agreement that sample matrix will often have a different consistency depending on freezing conditions. Based on these arguments, there did seem to be some support for validating stability at lower storage temperatures even if stability has already been determined at a higher temperature. Additional stability at lower temperature should be required for macromolecules and may also be performed for small molecules as needed.

Stability Criteria for Stock Solution Stability

The need to characterize the stability of stock solutions was emphasized throughout the meeting and accepted as a core validation experiment. However, there was no agreement on the degree of degradation that defines acceptable stability. The consensus was that lower degradation in the standard acceptable ranges is desirable since these stock solutions are used for making other solutions and this error may be propagated in the concentrations reported for biological samples.

Acceptance Criteria for Internal Standards

The practice of placing precision criteria on internal standards as an additional run acceptance test was discussed. A highly variable internal standard can be an indication of an uncontrolled process during sample analysis, especially if the internal standard response is variable with incurred samples. It is recognized that the internal standard is present in a bioanalytical assay to compensate for variability of extraction in LC-MS/MS analysis. This is most likely to be the case when the internal standard is isotopically labeled. When assessing the impact of internal standard variability, it is important to determine, in cases of low internal standard response, that the assay continues to have the ability to accurately quantify at the LLOQ. No agreement was reached on the inclusion of internal standard criteria or on the magnitude of acceptable internal standard precision. However, if study samples or analytical runs are rejected or repeated based on internal standard response variability, objective criteria are necessary and need to be established a priori.

CONCLUSION

For quantitative bioanalytical method validation procedure and requirements, there was a relatively good agreement between chromatographic assays and ligand binding assays. It was realized that the quantitative and qualitative aspects of bioanalytical method validation should be reviewed and applied appropriately.

1. Some of the major concerns between the two methodologies related to the acceptable total error for precision and accuracy determination and acceptance criteria for an analytical run. The acceptable total error for precision and accuracy for both the methodologies is <30 . The 4-6-15 rule for accepting an analytical run by a chromatographic method remained acceptable while a 4-6-20 rule was recommended for ligand binding methodology.
2. The Third AAPS/FDA Bioanalytical Workshop clarified the issues related to placement of QC samples, determination of matrix effect, stability considerations, use of internal standards and system suitability tests.
3. There was a major concern and issues raised with respect to stability and reproducibility of incurred samples. This should be addressed for all analytical methods employed. It was left to the investigators to use their scientific judgment to address the issue.
4. In general, the Third AAPS/FDA Bioanalytical Workshop provided a forum to discuss and clarify regulatory concerns regarding bioanalytical method validation issues.

ACKNOWLEDGEMENT

Contributors to this report include: Mark Arnold (Bristol-Myers Squibb Inc., Princeton, NJ 08543), Bruno Boulanger (Eli Lilly, Mont-Saint-Guibert, Belgium), Ronald Bowsher (LINCO Diagnostic Services, Saint Charles, MO 63304), Ajai Chaudhary (Lilly Research Laboratories, Indianapolis, IN 46285), Daksha Desai-Krieger (Johnson & Johnson Inc., Spring House, PA 19477), Binodh DeSilva (Amgen Inc., Thousand Oaks, CA 91320), John Findlay (Pfizer Global Research and Development, Groton, CT 06340), Boris Gorovits (Wyeth, Pearl River, NY 10965), Howard M. Hill (Huntingdon Life Sciences, Huntingdon, Cambridge, PE17 5HS, UK), Thomas Huggins (Procter & Gamble Pharmaceuticals, Mason, OH 45040), Christopher James (Amgen Inc.,

Thousand Oaks, CA 91320), John Kamerud (Lilly Research Laboratories, Indianapolis, IN 46285), Marian Kelley (Cencacor Inc., Wayne, PA 19087), Masood Khan (Covance Laboratories, Chantilly, VA 20151), Jean Lee (Amgen Inc., Thousand Oaks, CA 91320), Zhenmin Liang (Hoffmann-La Roche Inc., Nutley, NJ 07110), Steve Lowes (Advion Bio-Sciences Inc., Ithaca, NY 14850), Kamal Midha (University of Saskatchewan, Saskatoon, SK S7N-3R2 CANADA), Eric Ormsby (Health Canada, Ottawa, Ontario, CANADA), Marie Rock (Midwest BioResearch LLC, Oak Park, IL 60302), Bonita Rup (Wyeth Research, Andover, MA 01810), Michael Skelly (Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD 20857; see article footnote), Wendell Smith (Bowsher Brunelle and Smith, LLC, IN 46140), Jo Marie Smolec (Alta Analytical Laboratories Inc., San Diego, CA 92124), Richard Tacey (PPD Development, Richmond, VA 23233), Steve Unger (Bristol-Myers Squibb, New Brunswick, NJ 08903), Patrick Vallano (Mylan Pharmaceuticals Inc., Morgantown, WV 26505), Eric Woolf (Merck Research Labs, West Point, PA 19486), Jing-Tao Wu (Millennium Pharmaceuticals Inc., Cambridge, MA 02139), Martin Yau (Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD 20857; see article footnote).

REFERENCES

1. V. P. Shah, K. K. Midha, and S. V. Dighe. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies (conference report). *Eur. J. Drug Metab Pharmacokinet.* **16**(4): 249–255 (1992).
2. V. P. Shah, K. K. Midha, J. W. A. Findlay, H. M. Hill, J. D. Hulse, I. J. McGilveray, G. McKay, K. J. Miller, R. N. Patnaik, M. L. Powell, A. Tonelli, C. T. Viswanathan, and A. Yacobi. Bioanalytical method validation—a revisit with a decade of progress. *Pharm. Res.* **17**(12): 1551–1557 (2000).
3. Food and Drug Administration. *Guidance for Industry: Bioanalytical Method Validation*, US Department of Health and Human Services, FDA, Center for Drug Evaluation and Research, Rockville, MD, 2001.
4. 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).
5. Food and Drug Administration. *Draft Guidance for Industry: Safety Testing of Drug Metabolites*, US Department of Health and Human Services, FDA, Center for Drug Evaluation and Research, Rockville, MD, 2005.