The AAPS Journal 2007; 9 (1) Article 11 (http://www.aapsj.org).

Themed Issue: Bioanalytical Method Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays
Guest Editors - Mario L. Rocci Jr., Vinod P. Shah, Mark J. Rose, Jeffrey M. Sailstad

Key Elements of Bioanalytical Method Validation for Small Molecules

Submitted: February 7, 2007; Accepted: March 8, 2007; Published: March 30, 2007

Surendra Bansal[1] and Anthony DeStefano[2]

[1]Hoffman-LaRoche Inc, Nutley, NJ
[2]Procter & Gamble Pharmaceuticals, Mason, OH

Abstract

Method validation is a process that demonstrates that a method will successfully meet or exceed the minimum standards recommended in the Food and Drug Administration (FDA) guidance for accuracy, precision, selectivity, sensitivity, reproducibility, and stability. This article discusses the validation of bioanalytical methods for small molecules with emphasis on chromatographic techniques. We present current thinking on validation requirements as described in the current FDA Guidance and subsequent 2006 Bioanalytical Methods Validation Workshop white paper.

Keywords: bioanalytical, validation, precision, accuracy, sensitivity, selectivity, reproducibility

Introduction

Bioanalytical methods are used for the quantitation of drugs and their metabolites in biological matrices. In today's drug development environment, highly sensitive and selective methods are required to quantify drugs in matrices such as blood, plasma, serum, or urine. Chromatographic methods (high-performance liquid chromatography [HPLC] or gas chromatography [GC]) have been widely used for the bioanalysis of small molecules, with liquid chromatography coupled to triple quadrupole mass spectrometry (LC/MS/MS) being the single most commonly used technology. After developing a method with desired attributes, the method is validated to establish that it will continue to provide accurate, precise, and reproducible data during study-sample analysis. Method validation is a process that demonstrates that the method will successfully meet or exceed the minimum standards recommended in the Food and Drug Administration (FDA) Guidance[1] for accuracy, precision, selectivity, sensitivity, reproducibility, and stability. The validation is performed using a control matrix spiked with the compounds to be quantified. This article discusses the validation of bioanalytical methods for small molecules with emphasis on chromatographic techniques. We present current thinking on validation requirements as described in the current FDA Guidance[1] and subsequent 2006 Bioanalytical Methods Validation Workshop white paper.[2]

Corresponding Author: Surendra Bansal, 340 Kingsland Street, Nutley, NJ 07110. Tel: (973) 235-5919; Fax: (973) 235-7010; E-mail: Surendra.bansal@roche.com

Validation Parameters

Bioanalytical methods can be developed in the laboratory conducting the validation or obtained from another laboratory or literature. The results from a method validation can be no better than the quality of the method that was developed. Thus, before beginning the method validation, it is important that the method is set up and tested in the laboratory. For methods obtained externally, modifications may be necessary to achieve the desired performance of the method relative to how it was developed originally. This process will help ensure that when validation begins, chances for its successful completion (and more important, successful sample analysis) are high. During method validation, values for validation parameters are obtained. The essential parameters required according to the FDA Guidance[1] are selectivity, sensitivity, accuracy, precision, reproducibility, and stability. While obtaining these parameters, other parameters are also determined during validation (eg, extraction efficiency, calibration range and response function [linear or nonlinear], positional differences within an analytical run, and dilution integrity for analyzing above limit of quantitation [ALQ] samples). These validation parameters are described below in detail and are summarized in Table 1.

Selectivity

Selectivity or specificity should be assessed to show that the intended analytes are measured and that their quantitation is not affected by the presence of the biological matrix, known metabolites, degradation products, or co-administered drugs. Specificity should be determined for each analyte in the assay. Selectivity determination depends on the type of the assay as discussed below.

In assays wherein the intrinsic selectivity is low (eg, HPLC or GC with detection other than MS), it is necessary to confirm, using blank matrices from at least 6 independent sources, that the biological matrix will not interfere significantly with the assay. The same matrix as in samples should be used whenever possible. A proxy matrix is allowed if the sample matrix

Table 1. Synopsis of Validation Parameter Requirements*

Parameter or Process	Requirement
Selectivity (matrix interference)	Review noninterference in at least 6 sources of matrix for non-MS assays. For MS assays determine MFs in 6 sources if the nonisotopically labeled IS is used. If isotopically labeled IS is used, demonstrate that IS-normalized MF is close to unity.
Validation batches	Analyze at least 3 batches for accuracy and precision. At least 1 validation batch should be made as large as the largest anticipated sample analysis batch.
QC samples	Concentration of QC samples should be: Low QC: About 3 times the LLOQ Mid QC: Middle of the range (at about the geometric mean of low and high QC concentration) High QC: Near the high end of the range, ~70% to 85% of ULOQ Dilution QC: Sufficient to cover highest anticipated dilution
QC acceptance criteria	Intra- and inter-batch precision (%CV) and accuracy (%RE) should be: QCs prepared at all concentrations greater than LLOQ ≤15%; QC prepared at LLOQ concentration ≤20%
Calibration standards	Include the following calibration standards with each batch: Minimum of 6 non-zero standards Matrix blank: Matrix sample without internal standard Zero standard: Matrix sample with internal standard
Standard acceptance criteria	Acceptance criteria for calibration standards are: LLOQ standard ≤20% All other standards ≤15% At least 75% of standards should meet above criteria
Matrix blank	Interference in matrix blank should be ≤20% of LLOQ response
Recovery	Extent of recovery of analyte and IS should be consistent, precise, and reproducible. Determine recovery at 3 concentration levels.
Stability	Perform the following stability experiments: Stock solution: Minimum of 6 hours at room temperature Postpreparative (extracted samples/autosampler tray): Longest time from preparation through sample analysis. Assess against fresh standards, except for autosampler reinjection reproducibility. Benchtop: Stability at ambient temperature (or temperature used for processing of samples) to cover the duration of time taken to extract the samples (typically ~4-24 hours). Freeze-thaw: QC samples at minimum of 2 concentrations, 3 cycles, completely thawed, refrozen at least 12 hours between cycles, at anticipated temperature of sample storage. Long-term: Cover longest time from collection to final analysis for any sample in study. Analyze 3 aliquots at low and high concentrations with fresh standard curves and compare against intended (nominal) concentrations. Long-term stability can be completed postvalidation.

*MS indicates mass spectrometry; MF, matrix factor; IS, internal standard; QC, quality control; LLOQ, lower limit of quantitation; ULOQ, upper limit of quantitation; CV, coefficient of variation; and RE, relative error.

is of limited availability. The blank matrix should not produce any significant interference at the retention time of the analytes. For chromatographic assays, the peak response in the blank matrix at the retention time of analyte(s) should be no more than 20% of the response for the lower limit of quantitation (LLOQ) sample.

In assays where the selectivity is high (eg, LC-MS/MS or GC-MS/MS assays), it is less likely that co-eluting peaks will directly interfere with the quantification of the analytes. However, the presence of unmonitored, co-eluting compounds from the matrix may affect the detection of analytes. This phenomenon is commonly known as matrix effect. The whitepaper from the 3rd Bioanalytical Workshop[2] has proposed determination of matrix factors from 6 independent sources of matrix as a way of assessing the matrix effect. Matrix factor (MF) has been defined as

$$\text{Matrix Factor} = \frac{\text{Peak Response in Presence of Matrix Ions}}{\text{Peak Response in Absence of Matrix Ions}} \quad (1)$$

where peak response is defined as the peak area, peak height, peak area ratio (PAR), or peak height ratio (PHR) of chromatographic peaks. Peak area (or height) ratio is the ratio of the peak area (height) for the analyte vs that of the internal standard (IS).

The MFs can be determined for the analyte and the IS separately and a ratio of the 2 factors yields the IS-normalized MF for the analyte. The IS-normalized MF can also be determined directly by using peak response ratios (PAR or PHR) in the above equation. Because of the similarities in chemical properties and elution times of the stable-isotope labeled internal standards relative to the analytes, the MFs for an analyte and its stable-isotope labeled IS are usually similar. The IS-normalized MFs using stable-isotope labeled IS are therefore usually close to unity for bioanalytical samples. This has a very positive influence in reducing the variability of the assay due to matrix effects and makes the use of stable-isotope labeled internal standards very desirable in MS-based assays. It is recommended that matrix factors or IS-normalized MF be determined in 6 independent lots of matrices. The variability in matrix factors, as measured by the coefficient of variation (CV), should be less than 15%. When stable isotope labeled IS are used, it is not necessary to determine the IS-normalized MF in 6 lots. Determination of the IS-normalized MF in one lot is sufficient if its value is close to unity. If the value is not close to unity, use of the full 6 independent lots may be advisable.

Sensitivity

Sensitivity of the method is defined as the lowest concentration that can be measured with an acceptable limit of accuracy and precision. The accuracy and precision at the lower limit of quantitation (LLOQ) should be determined by analyzing at least 5 replicates of the sample at the LLOQ concentration on at least one of the validation days. These samples should be independent of those used for construction of the calibration curve. The accuracy as determined by the relative error (RE%) at this concentration should be within ±20% and the CV should be less than 20%. For this experiment, all results of sensitivity samples should be used to calculate accuracy and precision, including the values that fall below LLOQ.

Accuracy and Precision

Accuracy and precision of the assay should be determined for both intra- and inter-runs. They are determined by analyzing quality control (QC) samples at a minimum of 3 concentrations (low, mid, and high), representing the entire range of the calibration curve. The concentration of low QC should be near the lower limit of quantitation (no more than 3 times the LLOQ concentration). The mid-QC concentration should be somewhere in the middle of the calibration range. It is recommended that the mid-QC concentration be near the geometric mean of the low- and high-QC concentrations. The high-QC concentration should be near the upper end of the calibration curve (within the upper quartile of the calibration range). At least 5 replicates at each concentration should be analyzed. In addition to determining the accuracy and precision of these QC samples, the accuracy and precision at the LLOQ level should also be determined as described in the "Sensitivity" section. Although not required, accuracy and precision at upper limit of quantitation (ULOQ) may also be determined in a similar manner.

For intra-run accuracy and precision, the mean and CV of observed QC concentrations within a run should be determined. The mean of the observed concentrations should be within ±15% of the nominal at all concentrations of the QC samples. Coefficients of variation (indicating precision) around the mean observed concentration should not exceed 15% at all concentrations. For accuracy and precision, the mean and CV of the QC samples at each concentration from multiple runs (at least 3) should be determined. The mean observed concentration should be within ±15% of the nominal concentration, and the CV should be less than 15%, at all concentrations. If the QC concentration is at the lower limit of quantitation, the RE% and the CV can be up to 20%. For both intra- and inter-run precision and accuracy, all QC samples, including those that failed with no assignable cause, should be used for calculation. Only those QC samples that failed for an assignable cause (eg, rejected chromatography or sample extraction problem) should be excluded from the calculation of precision and accuracy.

Alternative methods for calculation of precision and accuracy (eg, analysis of variance [ANOVA]) are also acceptable. ANOVA procedures can be obtained from some popular commercial laboratory information management system (LIMS) software or from the AAPS Web site (www.aapspharmaceutica.org).

Reproducibility

Reproducibility of the assay is established during validation by the following tests:

- Precision and accuracy: Inter-run precision and accuracy is determined by analysis of the QC samples; see "Accuracy and Precision" section.
- Second column or instrument verification: To demonstrate the reproducibility of the method on an alternate column or instrument, it is recommended that a batch

of precision and accuracy samples is analyzed on a different column or instrument on one of the days of validation. This method is a good practice but is not required for all validations.

- Reproducibility using incurred samples: Reproducibility using incurred samples should be shown if samples are available. This test can be postponed and performed during sample analysis, where it is more important to prove the reproducibility of incurred samples analysis.

Stability

Several types of stability should be evaluated during the validation. Suggested experiments to determine stability are provided below. Alternate experiments that evaluate equivalent aspects of stability may be performed.

- Stock solution stability: The stability of the stock solutions of drug and internal standards should be evaluated at room temperature for at least 6 hours. If the stock solutions are kept refrigerated or frozen over a period of time, the stability over that period should be evaluated by comparing the response of the aged stock solution to that of a freshly prepared stock solution. Stock solution stability should be performed at one concentration in at least duplicate.
- Postpreparative (extracted samples/autosampler tray) stability: This stability is determined for ~48 to 96 hours to cover the anticipated run time for the analytical batch and to allow for delayed injection owing to unforeseen circumstances (eg, an instrument malfunction or the need to store samples over a weekend prior to analysis). The extracted QC samples (ready to inject) are kept at autosampler temperature for the established time and analyzed with fresh standards.
- Benchtop stability: Replicate (eg, triplicate) QC samples in matrix at a minimum of 2 concentrations are analyzed after keeping them at ambient temperature for 4 to 24 hours to cover at least the duration of time it takes to extract the samples. The observed sample concentrations are compared with their nominal values. This experiment can be combined with that for the extracted samples/autosampler tray stability above to demonstrate overall process stability, if desired.
- Freeze-thaw stability: QC samples in matrix at a minimum of 2 concentrations (eg, low and high QC concentrations) are frozen overnight, at normal storage temperature (eg, −20°C or −70°C) and thawed unassisted at room temperature. When completely thawed, the samples are frozen again at the same temperature for 12 to 24 hours and thawed. This freeze-thaw cycle is repeated 2 more times. After the third cycle, the samples are analyzed. The observed concentrations are compared with their nominal values. If an unacceptable level of degradation is observed, cycles 1 and 2 are repeated to determine where the instability occurs. The number of freeze-thaw cycles can be extended if needed.
- Freezer storage stability: During validation, stability at the nominal freezer storage temperature should be determined to the extent possible. However, longer term stability should be determined and appropriately documented, as discussed below.
- Postvalidation long-term stability: After validation is complete, long-term stability of the analyte(s) in the matrix should be determined by storing a sufficient number of QC samples at the required long-term storage temperature and analyzing them in at least triplicate at a minimum of 2 QC concentrations (eg, low and high QC concentrations). The long-term stability should be determined at several time points (eg, 1, 3, 6, 9, and 12 months) depending on the length of stability required. If possible, it is recommended that some stored in vivo samples are analyzed to assess the long-term stability of incurred samples at storage temperature. Upon obtaining the long-term stability data, the validation report can be amended to include the stability results or a separate report can be written to describe the long-term stability.

Extraction Efficiency (Recovery)

The extraction efficiency is a ratio of the detector response of an analyte from an extracted sample to the detector response of the analyte from an unextracted sample containing the same amount of analyte that was added to the extracted sample. The unextracted sample can be made up in solvents and is not taken through the extraction process. Alternatively, blank samples can be extracted and the extracts fortified with the analytes after extraction. These preparations represent 100% recovery during extraction. Extraction efficiency need not be very high, but it should be consistent, precise, and reproducible. Extraction efficiency can also be determined for the IS, and the ratio of the extraction efficiencies of the analyte and IS provides an IS-normalized extraction efficiency.

Calibration Range and Response Function

The relationship between the detector response and concentration should be demonstrated to be well defined and reproducible. A calibration curve should consist of a blank sample (matrix sample processed without the IS), a zero standard (matrix sample processed with internal standard), and 6 to 8

Table 2. Glossary of Common Bioanalytical Method Validation Terms. Definition of many of the terms given in the table are available in FDA guidance[1] or other publications, but are provided here for convenience

Accuracy	The degree of closeness of the observed concentration to the nominal or known true concentration. It is typically measured as relative error (%RE).
Precision	Measurement of scatter for the concentrations obtained for replicate samplings of a homogeneous sample. It is typically measured as coefficient of variation (%CV).
Selectivity	The ability of the bioanalytical method to measure and differentiate the analytes in the presence of components that may be expected to be present. These could include metabolites, impurities, degradants, or matrix components.
Sensitivity (LLOQ, lower limit of quantification)	The lowest concentration of an analyte in a sample that can be quantitatively determined with an acceptable precision and accuracy.
Sensitivity (ULOQ, upper limit of quantification)	The highest amount of an analyte in a sample that can be quantitatively determined with an acceptable precision and accuracy.
Standard curve	The relationship between the experimental response value and the analytical concentration.
Linearity	The ability of the bioanalytical procedure to obtain test results that are directly proportional to the concentration of analyte in the sample within the range of the standard curve.
Quantification range	The range of concentration, including the LLOQ and ULOQ that can be reliably and reproducibly quantified with suitable accuracy and precision through the use of a concentration-response relationship.
Recovery	The extraction efficiency of an analytical process, reported as a percentage of the known amount of an analyte carried through the sample extraction and processing steps of the method.
Matrix factor	A quantitative measure of the matrix effects due to suppression or enhancement of ionization in a mass spectrometric detector.
Stability	The chemical or physical stability of an analyte in a given matrix under specific conditions for given time intervals.
Reproducibility	Ability of the method to yield similar concentration for a sample when measured on different occasions.

nonzero standards. The number of standards can be increased for a complex curve or a curve covering a very large range. The simplest relationship that provides acceptable back-calculated concentrations for the standards should be used to fit the calibration curve. If a weighting factor is used, it should be defined during validation. The concentrations of calibration standards are back-calculated, and the residuals (difference between the back-calculated concentration of the calibration standard and its nominal concentration) determined. The residuals should be no more than ±15% at all concentrations except at the LLOQ level, where they can be up to ±20% of the nominal value. To accept an analytical run, at least 75% of the calibration standards should meet the stated acceptance criteria. Calibration standards not meeting the acceptance criteria should be eliminated from the calibration curve calculations. No extrapolation from the calibration curves is allowed, therefore the range of the calibration curve will be truncated if the end points on the calibration curve are eliminated.

Positional Differences

During a chromatographic analysis, samples are injected in sequence over several hours. Therefore, it is important to determine if the sample position in the chromatographic run sequence has an influence on the observed response (eg, if there is response change over the course of the run or any carryover is observed from previous samples). An evaluation of the situation should be done during the validation of the method and monitored during sample analysis. Procedures

for such determination may vary with the individual method or current practices observed in a laboratory. The following are some suggestions, which may be applied to assess the positional differences.

During validation of the method, analyze calibration standards and quality control samples in a predefined order (eg, as given below):

- Place 1 set of calibration standards at the beginning of the run (front curve).
- Analyze quality control and any other validation samples in the middle of the run, distributed randomly or placed in such a sequence as to help assess positional differences.
- Analyze 1 set of calibration standards at the end of the run (back curve).
- Place a blank matrix sample or zero standard after the high concentration sample to help assess carryover.

When using 2 sets of calibration standards, preferably place 1 set in ascending order and the other in descending order. After the analysis is complete, use both sets of calibration standards to construct the calibration curve. This calibration curve is used to determine the intra- and inter-run accuracy and precision. Alternatively, construct individual front and back calibration curves and determine the concentrations in quality control samples using the individual curves. Concentrations of the QC samples calculated by these 2 calibration curves will show a bias if there are positional differences. If there is no bias over a large number of samples in the analytical run, a single calibration curve may be used. The number of samples in at least 1 of the validation runs should approach the expected number of samples in a typical samples analysis run.

The peak response in the blank matrix sample or zero standard placed after the high concentration sample should be reviewed. The analyte response in this sample should be generally less than 20% of that of the LLOQ sample. If there is a known carryover in the assay that cannot be avoided, specific instructions should be provided in the method to deal with the carryover problem (eg, placing of blanks after expected high concentration samples).

Dilution Integrity for Analyzing Above Limit of Quantitation Samples

If it is expected that some sample concentrations may exceed the upper limit of quantitation, a test for sample dilution with blank matrix during validation should be performed. One or more additional QC samples at concentrations several times higher than the upper limit of the calibration curve should be prepared, covering the maximum expected dilution. These QC samples are diluted with blank matrix to bring the concentration to within the calibration range and then analyzed. The acceptance criteria for the diluted QC are the same as provided in the "Accuracy and Precision" section. Dilution integrity is performed on at least 1 day of validation. If during sample analysis a dilution higher than the one covered during validation is needed, further dilution can be validated during samples analysis by analyzing the required diluted QC samples.

Common Terminology Used in Bioanalytical Methods Validation

Table 2 contains a glossary of common terms used in bioanalytical methods validation.

References

1. Food and Drug Administration. *FDA Guidance for Industry: Bioanalytical Method Validation.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2001.

2. Viswanathan CT, Bansal S, Booth B, et al. Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *AAPS J*. 2007;9:E30-E42.