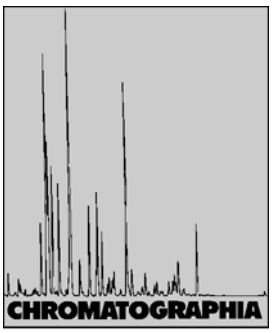
Validation Issues Arising from the New FDA Guidance for Industry on Bioanalytical Method Validation



2002, 55, Suppl., S-91–S-94

R. R. Brodie* / H. M. Hill

Huntingdon Life Sciences, Alconbury, Cambridgeshire, PE28 4HS, UK; E-Mail: BrodieR@ukorg.Huntingdon.com

Key Words

Hyphenated methods
Bioanalytical method validation
FDA Guidance

Summary

In late May 2001 the FDA issued their final Guidance for Industry on Bioanalytical Methods Validation. Different types and levels of validation, dependent on the status or changes to the validated method, are now defined and characterised. Additional experiments with the emphasis on the newer hyphenated analytical techniques and changes to quality control and calibration acceptance criteria are now required. A number of inconsistencies, in terms of the minimum number of calibration points required to define a calibration curve, exist in the current document. It is therefore important to understand the underlying philosophy of this guidance and what are the most appropriate ways of implementing this philosophy.

Introduction

The current guidelines have been developed from a culmination of a number of meetings and publications during a ten-year period.

- Consensus Meeting, December 1990, Arlington, Washington.
- Consensus Document published April 1992 [1].
- FDA Draft Guidance, "Human Studies", December 1998.
- Bioval '99, June 1999, London
- Consensus Meeting, January 2000, Arlington, Washington (small molecules)
- Consensus Meeting, March 2000, Arlington, Washington (macromolecules)
- Summary Over View – circulated to Attendees
- Consensus Document published December 2000 (March 2001) [2].
- FDA Guidance for Industry, Bioanalytical Method Validation, May 2001

The scope of the final guidance document encompasses the quantification of "small molecules" in biological matrices from all species for the purposes of kinetic or pharmacodynamic evaluation. The original draft document was some 10 pages long – and was limited to human studies. The new document is now 22 pages long and encompasses all species, all matrices and all methods, be they "chemical" or ligand-based assays. A benefit of the latest guidance is the standardisation in terminology and defining the different types of method validation and regulatory requirements.

Discussion

Validation Definitions

Different types and levels of validation are defined and characterised as follows:

Full validation is required for a new drug entity and when implementing a developed method for the first time. Full validation is also a requirement if metabolite measurements are added to an existing method.

Partial validations are required if modifications are made to an already validated method. These validations will vary depending on the extent of the modification and the impact on the integrity of the validation data. The experiments required for a partial validation will vary from as little as a one-batch intra-assay precision and accuracy determination, for example when transferring a method between analysts, to almost a full validation, for example when there is a change in species within a matrix (rat plasma to mouse plasma).

Presented at: 14th International Bioanalytical Forum: Sensitive Bioanalysis in Anti-cancer and other Drug Areas, Guildford, UK, Jul 3–6, 2001

cf. Hill's report after 13th Forum, p. S-65 in Supplement to Vol. 52 (2000).

0009-5893/00/02 S-91-04 $ 03.00/0

Cross-validation is defined as a comparison of two or more different bioanalytical methods (e. g. LC-MS-MS vs HPLC) or a comparison between sites or laboratories, a requirement if data within a single study is being generated at different sites or laboratories. Cross-validation consists of the analysis of both spiked quality control samples and "real" test samples, analysed both ways (by both methods or at both sites/laboratories).

The Validation Process

The guidance sub-divides the validation process into 3 distinct phases, namely the reference standard phase, the development/establishment phase and the application phase.

For a validated method it is a requirement that the reference standard should be identical to the analyte being measured or an established chemical form of it (i. e. free base or acid, salt or ester). The standard should be authenticated in terms of identity and purity and stored under conditions to maintain its integrity. The source, lot number, expiration date and ideally certificates of analyses should be available for each reference standard.

Once a method has been developed and established the fundamental parameters to be determined for a bioanalytical method validation are defined as accuracy, precision, response function (calibration curve), selectivity, sensitivity, reproducibility and stability for each analyte to be measured. Recovery of an analyte from the biological matrix should be optimised during the developmental phase, although recovery need not be 100% and no set criteria are required other than the extent of the recovery should be consistent, precise and reproducible. It is recommended that recovery experiments be performed across the calibration range (at low, medium and high concentrations).

Once a method has been validated and is applied to routine drug analysis, calibration measurements as well as accuracy and precision experiments should be undertaken with each batch of samples to ensure on-going validity of the method. In addition, estimation of concentrations above the highest calibration standard (upper limit of quantification, ULOQ) or below the lower limit of quantification (LLOQ) by extrapolation is not recommended. Samples should be analysed within their known stability period, and where appropriate (i. e. clinical, pharmacokinetic studies) all samples from the same subject, in any one phase of the study, analysed within a single batch.

Specific Method Validation Recommendations

Bioanalytical method validation includes all the experimental procedures and documentation which demonstrates that a particular method used for quantitative measurement of analytes is reliable and suitable for the intended analytical applications. Fundamental parameters that require determination are accuracy, precision, selectivity, sensitivity, reproducibility and stability. When considering the requirements for validation, in terms of suitability for use, it is important to understand the significance of the how analytical data impacts on the overall interpretation and objectives of the study. For example, validation requirements, in terms of precision, accuracy, specificity, etc. may be significantly different for methods required to analyse samples from human bioequivalence studies, where biological variability is less (serial sampling for the same subject, cross-over designs) compared to that of animal toxicokinetic studies, or analysis of tissue samples, where biological variation will be much greater (single sampling from individual animals),

From the guidance, the specific recommendations for method validation for chemical assays are as follows:

Matrix-based calibration/standard curves for each analyte should consist of a blank sample (matrix sample without internal standard), a zero sample (matrix sample spiked with internal standard) and six to eight non-zero standard points (concentrations), covering the entire concentration range and including the LLOQ. Throughout the guidance various numbers of standard points required to define a standard curve are quoted. In principle, additional points, ($n > 6$) should be included, particularly for non-linear relationships. The simplest model that adequately defines the concentration-response relationship should be used and the *goodness of fit* in terms of back-calculated responses of the individual concentrations should not deviate by more than 15% from the nominal concentration (20% at the LLOQ). The selection of weighting and the use of a more complex regression model should be justified. At least four out of six non-zero standards should meet the acceptance criteria, including the LLOQ and the highest calibration standard. Excluding standards should not change the model used. If the number of standard concentrations is greater than six, 75% or a minimum of six non-zero standards should be acceptable.

Accuracy as determined by replicate analysis of spiked samples containing known amounts of analyte should be measured at minimally four concentrations over the entire calibration range (low (up to $3 \times$ LLOQ), middle and high concentrations), as well as at the LLOQ of the method. A minimum of five determinations per concentration is required during a single analytical batch to establish within-batch accuracy. Inter-batch measurements should be determined by analysis of QC samples at the same concentrations on separate occasions (minimally three). The mean measured concentration should be within 15% of the actual concentration except at the LLOQ, where it should not deviate by more than 20%. The deviation of the mean from the true value is a measure of accuracy.

Precision, in terms of the closeness of agreement between measurements from multiple sampling of the same homogeneous sample, should be determined at a minimum of three concentrations and at the LLOQ of the method, at the concentrations defined for accuracy determination. A minimum of five determinations at each concentration within a single batch should be undertaken to establish within-batch precision. Inter-batch precision or repeatability, measuring precision over time (minimally three occasions), which may involve different analysts, equipment, reagents and laboratories, if appropriate, should be determined. The precision determined at each concentration should not exceed 15% coefficient of variation (CV), except at the LLOQ, where it should not exceed 20%.

Selectivity from potentially interfering substances (endogenous compounds, metabolites, decomposition products, and concomitant medication) should be established by the analysis of blank samples of the appropriate biological matrix obtained from at least six different sources. Each sample should be checked for interference to ensure selectivity at the LLOQ of the method for each analyte. No absolute criteria are set, except that the analyte response at the LLOQ should ideally be at least 5 times the response compared to the

blank response. In the case of hyphenated mass spectrometry-based methods, testing for interference may be less important; however matrix effects which may compromise the ionisation of the analyte should be investigated to ensure that precision, selectivity and sensitivity are not compromised.

Stability of each analyte in biological matrix should be confirmed, minimally at low and high concentrations (at least three replicates at each concentration) over the calibration range, during short-term storage at room temperature. The storage time should be based on the expected duration that test samples are maintained at this temperature during the intended study (4 to 24 hours). Stability should also be determined after three freeze and thaw cycles and after longer-term storage in the freezer, on three separate occasions, using identical storage conditions to those of the test samples (i. e. −20 °C or −70 °C). Long term storage should confirm stability over the longest period which test samples are stored, ideally prior to analysis.

Stability of each analyte and internal standard stock solution, if appropriate, should be evaluated at room temperature for at least six hours, and if refrigerated or frozen, during a relevant storage period. Stability should be tested by comparing the instrument response of the old solution with that of a freshly prepared solution. Once stability has been confirmed and accuracy of preparation verified, standards and QCs can be prepared from the same spiking stock solution.

The stability of processed samples, including residence time in the autosampler should be determined for the length of the anticipated batch size and to cover re-injection of samples, if appropriate. For labile analytes, investigations may include analysis of samples from dosed subjects.

The guidance states that concentrations of all stability samples should be compared to the mean of the back-calculated values for the samples, at the appropriate concentrations, measured on the first day of testing. The disadvantage of using this approach, instead of comparing measured values with the theoretical (nominal) spiked concentration, is that error measurements (which can be as much as ± 15% for each measurement) may make interpretation difficult. Generally speaking, if stability is a problem then a loss of compound with time should be evident (i. e. progressively decrease in concentration). When using a comparison with theoretical concentrations, stability will only become evident when accuracy measurements exceed the 15% error criteria.

The ability to dilute samples whose concentrations are above the upper limit of quantification (ULOQ) of the method both accurately and precisely should be demonstrated. This can be achieved by preparing a QC sample at a concentration higher than the anticipated maximum concentration expected in test samples (i. e. 10 × mid-point of calibration line) and diluting this sample 10-fold with the same biological matrix prior to analysis. The precision and accuracy of the measurement of this sample should meet the criteria previously discussed. Once the dilution has been validated the need to incorporate actual within-study over-range QC samples is obviated.

Special Issues Associated with Ligand-binding and Microbiological Assays

Many of the validation parameters discussed for chemical assays are also applicable to ligand-binding and microbiological assays. However, due to their unique characteristics other considerations should be taken into account during method validation.

Selectivity in terms of interference from substances with similar physiochemical properties to the analyte or matrix effects of unrelated compounds should be investigated. This should consist of checking for cross-reactivity of metabolites, endogenous compounds or concomitant medications, if appropriate. Interference should be checked, if possible, by assessing dilution linearity in test (incurred) samples or analysing these samples with an alternative reference method (LC-MS). Matrix effects should be tested by comparing responses of a standard curve prepared in buffer with that prepared in matrix. Also, a check for parallelism by diluting study samples and standards should be undertaken.

For quantification, as calibration curves are inherently non-linear, they should consist of a minimum of six non-zero concentrations, analysed in duplicate. In addition, anchoring points at the low and high ends of the calibration curve may improve the overall fit of the curve and therefore improve the accuracy of measurement, which is the key requirement. Due to the non-linear nature of the calibration curves both LLOQ and ULOQ need to be defined in terms of acceptable precision and accuracy based on study requirements. Generally, mean values should be within ± 20% of the theoretical and precision around the mean should not exceed 20% CV at these limits. Batch acceptance criteria are the same as for chemical assays; however, the guidance indicates that in certain situations wider criteria may be justified.

Application of the Validated Method

Once the method has been validated for routine use, its precision and accuracy should be monitored regularly to ensure continued performance. Matrix-based calibration curves, consisting of a minimum of six non-zero standard points (either single or replicate) and quality control (QC) samples at a minimum of three concentrations (low, middle and high), at least in duplicate, over the calibration range should be analysed with each batch of test samples.

The response function of the calibration curve, in terms of curve fitting, weighting and goodness of fit, should be the same as used during the validation phase. System suitability is used to ensure optimum operation of the analytical system.

Results of the matrix-based QC samples are the basis of accepting or rejecting batches. For a batch to be accepted, at least 67% (four out of six) of the QC samples should be within 15% of their respective theoretical values. This means that 33% of the QC samples may be outside the ± 15% criteria, but there must be at least one QC at each concentration level within 15% of its theoretical value. The minimum number of QC samples (in multiples of three) analysed in a batch should represent 5% of the total number of unknown samples or a minimum of six, whichever is the greater.

Estimations above the ULOQ or below the LLOQ are not recommended. High level samples should be diluted with matrix to fall in the validated calibration range.

In the case of multiple analytes, data from only one analyte failing the acceptance criteria should not preclude acceptance of data for the other analytes, which are acceptable.

Where samples need to be re-analysed or re-integrated, a guideline or standard operating procedure (SOP) should be in place to explain the rationale for these procedures.

Study Documentation

All validations and sample analyses should adhere to FDA Good Laboratory Practices (GLPs). General and specific SOPs and good record keeping are essential to support regulatory submissions.

Documentation requirements for method establishment and validation should provide a detailed operational description of the analytical method, including the purity and identity of the reference standards (compound, metabolites, internal standard, if appropriate) used. In addition, a description of all validation experiments and the relevant data obtained in these studies are required and this also includes stability studies.

Documentation should include legible examples of annotated chromatograms or mass spectrograms, if appropriate. Any deviations from SOPs, protocols and GLPs should be statet where applicable, and justifications for deviations.

Documentation to support the application of the validated methods should again include the purity and identity of the reference standards (compound, metabolites, internal standard, if appropriate) used. Chain of custody of the samples, in terms of sample identification, collection dates and times, storage conditions prior to and after shipment and their condition and storage prior to analysis should be documented and tabulated.

Summary tables of analytical batches should include batch (run) identification, date and time of analysis, method, analyst, start and stop times, duration, significant equipment and material changes, issues or deviations from the established method.

All calibration curve data, including equations used for back-calculation of results, QC sample summary and data on inter-assay accuracy and precision from calibration curves and QC samples used for accepting analytical batches, should be available.

Representative complete serial chromatograms of test samples including standards and QC samples, representing 20% of subjects for pivotal bioequivalence studies, are required. In other studies, 5% of randomly selected subjects in each study should be included. The selected chromatograms should be defined prior to sample analysis.

Reasons for missing samples, repeat analysis of samples and re-integration of data should be documented. Information should include the initial and repeated result, reason for the repeat, requestor and authoriser for the re-analysis. Repeat analysis and re-integration should be undertaken using predefined SOPs.

All deviations from the analysis protocol or SOPs, with reasons and justification, should be documented.

Conclusions

The latest Guidance has many similarities to the published proceedings [1, 2] from the Consensus meetings held during 1990 and 2000. However, significant changes, are now evident in this document with respect to minimum numbers of calibration points defining acceptable calibration curves and tighter QC batch acceptance criteria. The impact on these changes will undoubtedly be that more sample batches will fail and more sample re-analysis will be required, impacting on the timings and cost of studies. In addition, documentation requirements, as defined, will have the biggest impact and may require significant changes in working practices and SOPs.

References

[1] Shah, V.P.; Kamal, K.M.; Shrikant, D.; McGilveray, I.J.; Skelly, J.P.; Yacobi, A.; Layloff, T.; Viswanathan, C.T.; Cook, C.E.; McDowall, R.D.; Pittman, K.A.; Spector, S. *J. Pharm. Sci.* **1992**, *81*(3), 309–312.

[2] Shah, V.P.; Kamal, K.M.; Findlay, J.W.; Hill, H.M.; Hulse, J.D.; McGilveray, I.J.; McKay, G.; Miller, K.J.; Patnaik, R.N.; Powell, M.L.; Tonelli, A.; Viswanathan, C.T.; Yacobi, A. *Pharm. Res.* **2000**, *17*(12), 1551–1557.

Received: Oct 17, 2001
Revised manuscript received: Jan 15, 2002
Accepted: Jan 17, 2002

Addendum (summary by H. M. Hill):

Discussion at Forum of this and preceding presentations

Questions (D. Dell). – #Does the Method Validation Guidance cover Phase I studies as well as Bioequivalence studies? #What constitutes a 'macromolecule'? *Expressed opinions.* – #The guidance covers all bioanalytical applications even though only bioequivalence studies have a specific Guidance (21CFR 320) pertaining to them. #A biopolymer is a macromolecule if its MW exceeds 1000Da, its structure is difficult to define using conventional analytical techniques, and it has the potential to provoke an immune response.

C.A, James asked what would happen if a method could not meet the 15% acceptance criteria. *In reply, Hill* considered that if all avenues could be shown to have been explored (including replicate analysis) then 'broader' limits would be acceptable providing the assay could be shown to be in control throughout its application. In a similar vein, *D. Muirhead asked* if the Guidance provided a tick box approach to validation. *Hill's reply:* no! – the Guidance is indeed just that, i.e. a guidance. *Query by P. Heizmann:* need one randomly repeat 15% of the samples? *Opinion:* in the context of the FDA Guidance this is not considered to be a requirement, although Canadian studies still appear to need it.

W. Wild: if the LLOQ and ULOQ samples were rejected from the calibration, would the whole run need to be repeated? *Hill's view:* if in all other respects the 'run/batch' was acceptable, then data falling within the truncated curve should be acceptable. *Remark by J. Burrows.*– Guidance has an acceptance criterion that blank samples must not have an interference of greater than 20% of the LLOQ; but this acceptance is not contained within the routine batch acceptance criteria.

Dell: there is little reference to the need for cross validation between labs and related acceptance criteria. *Hill (cf. foregoing article):* a paragraph has been dedicated to this aspect of validation, although no clear acceptance criteria are provided.

Presentations by D. Browne (p. S-75) and R.D. McDowall (p. S-85). – *J. Schmid* wondered about validation of the smoothing processes in LC-MS-MS systems. *McDowall's response:* different smoothing factors are used depending upon the required sensitivity.

Editorial note (E. Reid). – *Much pertinent material appears elsewhere in this Volume, including literature cited in Appendix I. Entries worth consulting in the* General Index: – Calibration. Data ... LLOQ. Standards. Validation.