ELSEVIER

Journal of Chromatography B, 686 (1996) 3–10

JOURNAL OF CHROMATOGRAPHY B: BIOMEDICAL APPLICATIONS

Validation of liquid chromatographic and gas chromatographic methods
Applications to pharmacokinetics

Françoise Bressolle[a,*], Maguy Bromet-Petit[b], Michel Audran[c]

[a]Laboratoire de Pharmacocinétique, Faculté de Pharmacie, Université Montpellier I, 15 Avenue Ch. Flahault, 34060 Montpellier Cedex 01, France
[b]Biotec Centre, Orléans, France
[c]Laboratoire de Biophysique, Faculté de Pharmacie, Université Montpellier I, Montpellier, France

Abstract

Validations of analytical methods are important for the generation of data for bioavailability, bioequivalence and pharmacokinetic studies. It is essential to use well defined and fully validated analytical methods to obtain reliable results that can be satisfactorily interpreted. This manuscript is intended to provide guiding principles for the evaluation of a method's overall performance. For this purpose, all of the variables of the method are considered, including sampling procedure, sample preparation, chromatographic separation, detection and data evaluation. The criteria considered are as follows; stability, selectivity, limits of quantification and of detection, accuracy, precision, linearity, recovery and ruggedness. Models used for analytical calibration curves are explained in term of validity and limitations, along with a presentation of the most common statistical considerations used to validate the model. Appropriate means of testing precision and accuracy, the most important factors in assessing method quality, are presented. Other issues, such as re-validation, cross-validation, partial sample volume, endogenous drugs and biological matrix of limited availability, are also discussed.

Keywords: Pharmacokinetics

1. Introduction

Analytical methods for the quantification of drugs and their metabolites in biological samples play a significant role in the evaluation and interpretation of bioavailability, bioequivalence and pharmacokinetic data. The validation of bioanalytical methods has been the subject of discussion in recent conferences and papers [1–15].

The ultimate objective of the method validation process is to provide evidence that the method does what it is intended to do [1]. All of the variables of the method should be considered, including sampling procedure, sample preparation, chromatographic separation, detection and data evaluation, and the use of the same matrix as that of the intended samples.

The validation procedure includes first the validation of the analytical method [i.e., selectivity, precision, accuracy, recovery, limit of quantification (LOQ), limit of detection (LOD) and stability], and second the validation of the stability in the biological matrix. This validation procedure needs to be performed prior to the routine use of the analytical procedure.

This manuscript describes the characteristics that must be considered during validation of the ana-

*Corresponding author.

0378-4347/96/$15.00 © 1996 Elsevier Science All rights reserved
PII S0378-4347(96)00088-6

lytical procedures intended for registration applications submitted within the EC, Japan and USA.

2. Validation methodology

Method validation includes all of the procedures required to demonstrate that a method to quantify the concentration of an analyte (or series of analytes) in a particular biological matrix is reliable for the intended application.

2.1. Specificity/selectivity

The terms selectivity and specificity are often used interchangeably. Specificity is the ability to assess unequivocally the analyte in the presence of endogenous compounds. Selectivity includes the ability to separate the analyte from degradation products, metabolites and co-administered drugs.

There are a variety of ways to validate selectivity. Karnes et al. [8] suggest that the simplest test for chromatographic analysis is to demonstrate a lack of response in the blank biological matrix. Careful examination of blank chromatograms from several sources across the time windows of peaks of interest (i.e., comparison of retention times of endogenous compounds of the matrix to those of the drugs to be analysed) is required to evaluate specificity. Another approach suggested by the same author in the case of unweighted linear regression is to test whether the intercept of the calibration is significantly different from zero (one-sided *t*-test). Shah et al. [9,11] recommend establishing the specificity/selectivity of biological samples using six independent sources of the same matrix.

Real problems arise when analyte metabolites or known degradation products are not available. In this case, the most difficult task is to identify whether the peaks are pure within a sample chromatogram. Different ways should be used to check the selectivity of the method:

1. use of diode array or multiple-wavelength detectors to ensure peak purity,
2. use of a more specific detector, such as a mass spectrometer,
3. run multi-dimensional chromatography, i.e. the use of an analytical column with a different selectivity could be an alternative.
4. use of biological samples from dosed subjects. These samples should be analysed under different chromatographic conditions to resolve as many potentially merged peaks as possible. Moreover, examination of chromatograms from subjects' samples that were collected at various times following the drug intake can reveal peaks due to substances that are absent in the pre-dose sample.
5. use of standard additions of known quantities of analyte to real sample (from clinical trial for example) which may contain metabolites. The linear relationship between added analyte and response should be verified.

2.2. Calibration curve

It is necessary to use a sufficient number of standards to define adequately the relationship between concentration and response. The calibration curve should be constructed using at least five to eight values (excluding blank values) from the expected range of concentrations. Each standard is prepared by adding an appropriate volume of stock solutions; this volume should be always smaller than, or equal to, 2% of the total volume of the samples. Although some analytical procedures may require the use of non-linear calibration, it is customary to use a linear model, with the standard parameter estimation procedure based on the "least squares" methodology. In this approach, the independent variable (x) is concentration, the dependent variable (y) is response, and the computation procedure implicitly assumes that the measurement error is the same and normally distributed for each sample (y). Obviously, a different regression analysis approach has to be applied (i.e., extended or weighted least squares analysis with appropriate error model variance assumption) if the previously mentioned assumption is not applicable. In any case, the assumption concerning the measurement error must be verified at the end of the calculation procedure to validate the results found. This verification is mainly based on the analysis of the distribution properties of the residuals (i.e., the difference between observation and prediction)

which is expected to be normally distributed and centred around zero (Kolmogorov–Smirnov test). If the results found cannot support this assumption, the conclusions are that either the model for the calibration is incorrect, and/or the measurement error assumptions are inconsistent with the data. In both cases, the estimated parameters cannot be used. A better fit must be provided using a different error model, or, in some cases, using a non-linear model. In the latter case, more standard concentrations may be necessary to properly define the non-linear relationships than would be required for the linear one.

The quality of fit can be evaluated by comparing calculated standard points to the nominal ones. A linear regression of the calculated concentrations *versus* the nominal ones should provide a unit slope and an intercept equal to 0 (Student's t-test). The acceptability of the calculated mean for each calibration point should be in line with the acceptance criteria set for the evaluation of quality control (QC) samples.

Linearity of the method should be demonstrated by showing that the slope of the linear calibration curve is statistically different from 0, the intercept is not statistically different from 0 and that the regression coefficient is not statistically different from 1. If a significant non-zero intercept is obtained, it should be demonstrated that there is no effect on the accuracy of the method.

Bioanalytical laboratories react differently, and somewhat arbitrarily, regarding inclusion or exclusion of a standard point deviating greatly from the calibration curve. Some leave it in the curve and some drop it. Provided that the calibration curve consists of at least seven non-zero single standards, up to two non-zero standards may be removed from the calibration curve if at least one of the following valid reasons exists and a minimum of five non-zero standards remain in the curve:

1. loss of sensitivity,
2. poor chromatography,
3. losses during sample processing,
4. if, when included in the calibration curve, it clearly biases the QC results, and the back-calculated standard concentration deviates substantially from its nominal value.

In order to generate an accurate "analytical" calibration curve independent of the possible time effect, prepared freshly while QC samples are prepared and stored frozen at the same temperature as is intended for storage of study samples in order to account for the "time effect". A simple approach is to prepare a series of working calibration standards (in purified water for example) at concentrations that are ten to twenty times higher than those intended for biological standards. These working calibration standards may be stored (at 4°C or −20°C) provided that their stability has been demonstrated previously over the maximum period over which they will be stored. Then, on a daily basis, blank biological matrix is spiked with the working calibration standards at a ratio of e.g. 1:20 working standard–biological blank. To compensate for the dilution of biological matrix with working standards, an equal volume of working solution, free of analyte, is added to the study samples. However, in some analytical laboratories, the calibration curves are prepared and frozen for storage with QC samples.

2.3. Limit of quantification

There is a great deal of confusion over the terms related to the ability to assay low concentrations. A term frequently used is sensitivity. A method is said to be sensitive if small changes in concentration cause large changes in analytical response [16]. The ability to detect small concentrations is expressed as the LOD or LOQ.

The LOQ must be differentiated from the LOD. LOD is the smallest concentration that can be distinguished from the noise level. The LOQ should be at least twice the response of the LOD. Inter- and intra-day precision and accuracy of the LOQ can be determined by using at least five QC samples, from a single pool of matrix, independent of standards. The mean values should be within pre-defined boundaries, normally within ±20% of the nominal concentration, with a coefficient of variation ≤20%. The LOQ should serve as the lowest concentration on the calibration curve.

Concentrations below the LOQ should not be quantitatively reported, but may be reported as being simply "present" or as semiquantitative numbers.

They should not be used for the interpretation of results without appropriate weighting factors.

2.4. Precision and accuracy

Precision and accuracy together determine the error of an analytical measurement and are the primary criteria used when one judges the "quality" of an analytical method. Precision and accuracy are often expressed relative to one day or relative to a period of days

2.4.1. Quality control samples

After checking the specificity of the method, an adequate volume of biological matrix should be reserved and used during method validation and during the pharmacokinetic study.

Sufficient quantities of each QC concentration are prepared. A minimum of three concentrations, representing the entire range of the calibration curve, should be studied; one near the LOQ, one near the centre and one near the upper boundary of the standard curve. These control batches are separated into aliquots, frozen in appropriate containers, and used in the method validation and subsequent pharmacokinetic study analyses.

The QC samples, that had been divided into aliquots, can be randomised to be analysed in numerical order and included in each analytical sequence during the study period.

2.4.2. Precision and accuracy determination

Precision is usually assessed on both a within-batch and a between-batch basis, with this terminology being more appropriate than "within-day" and "between-day". Between-batch assessment is not always carried out with a single batch per day, and some batches may be of sufficient size that more than one day is required for analysis. Within-batch assessment should be considered as a measure of the precision of a method under optimal conditions. The between-day batch precision is considered to be a better representation of the precision one might observe during routine conduct of a method, because these data are generally subjected to a greater source of variability.

The accuracy and precision should be determined with a minimum of five determinations per QC sample (excluding blank matrix) from an equivalent biological matrix. The precision around the mean value should not exceed 15% of the coefficient of variation and the mean value should be within ±15% deviation of the nominal value for accuracy. It is desirable that these tolerances be provided both for intra-day and inter-day experiments. At the LOQ, 20% is acceptable for both precision and accuracy. The significance of the 15 and 20% limits has been discussed by Hartmann et al. [14]. It is possible to test simultaneously inter- and intra-day precision and accuracy, by performing replicate analyses of each concentration per run over several runs. Three day experiments are usually done. In this case, the residual variability should be ≤15% (ANOVA with repeated measurements).

The results of QC performed during sample analysis (i.e., pharmacokinetic study) provide the basis for accepting or rejecting the run. Indeed, values that fall outside a set rejection range should be deemed "out of control" and samples corresponding to these controls should be reanalysed. For each individual QC, the acceptance criterion is not more than a 20% deviation from the nominal value for accuracy.

2.5. Recovery

The recovery should be documented throughout the standard curve range. Absolute recovery is measured as the response of a processed spiked matrix standard, expressed as a percentage of the response of pure standard which has not been subjected to sample pre-treatment.

In order to study the effect of co-extracted biological material, recovery is computed by comparing responses of replicates of extracted QC samples with those of extracted blank matrix to which analyte has been added at the same nominal concentration.

If an internal standard is used, its recovery should be determined independently at the concentration level used in the method.

Values for recovery of not less than 50, 80 and 90% have all been used as numerical acceptance limits [8]. Although it is desirable that recovery has to be the highest possible, it is not needed to provide good accuracy and precision if adequate detection can be attained.

2.6. Stability

Stability data are based on duplicate or triplicate determinations of QC samples at two or three concentration levels (low, medium and high) at multiple time points after the start of storage to allow "trends" to be detected. However, the issue is not whether there is a trend in degradation, but whether the study samples are adequately preserved at the time of analysis.

2.6.1. Validation of the analytical method

First of all, stability of pure analyte and/or solutions of the analyte must be studied in replicate ($n \geq 3$), under normal laboratory conditions of heat, humidity, light and air exposure, by comparison with fresh solutions.

Method development should also investigate the stability of the analytes prior to chromatographic investigations, for example, during extraction, clean-up, phase transfer and during the storage of dry-residues or the supernatant from the extraction (in refrigerator or at −20°C) or of reconstituted extracts (in refrigerator or in autosampler). Each step of this stability study should be performed in replicate on QC samples, as indicated above. The mean value obtained after storage should be within ±5% of the nominal value (±15% if derivatisation has been carried out). These tests should be performed for the analyte(s) and the internal or external standard.

2.6.2. Validation of the biological sample

Two essential types of stability studies should be performed; (1) short-term stability, including bench-top storage, 4°C, −20°C and stability during freeze–thaw cycles and (2) long-term stability. Long-term stability must be proven over at least the maximum period of storage of study samples under the temperature conditions to be used for study samples (i.e., −20°C, −80°C) and using the exact type of container (e.g., glass, polypropylene) as used for the study samples. Generally, the assessment of long-term stability of frozen samples is achieved by comparing the responses of frozen and freshly prepared QC samples. Dadgar et al. [15] propose a more precise and accurate method to study long-term stability: QC samples are stored both at the temperature intended for use and in liquid nitrogen, the latter being the reference samples. However, the use of liquid nitrogen equipment is a limiting factor in the application of this method.

Stability in blood samples could be checked in order to validate the blood sample procedure.

Ideally, part of the control samples used for the evaluation of long-term stability should be obtained from dosed patients or from volunteers and should be stored at the same time and under the same conditions as the spiked samples. This provides maximum assurance of the integrity of all analytes in the study samples for a given study. For example, such a procedure should detect hydrolysis of glucuronide during storage. However, blank matrix samples spiked at different concentrations are an acceptable alternative and are usually used.

2.7. Ruggedness

The ruggedness of a method can be assessed by studying the eventual effect of different sets of conditions on the method. This is done through cross-validation. Typical reasons for cross-validation should include transfer of the method from one analyst to another, significant instrumental or procedural modifications (in HPLC, the difference in chromatographic performance between columns of the same designation is the most common source of chromatographic variability) and a significant time lapse between periods of operation. Cross-validation will involve the regeneration of validation under the new conditions and comparison of these results with the original.

An *a posteriori* assessment of ruggedness under different conditions is necessary when one wishes to carry out inter-laboratory comparisons.

3. Biological matrix of limited availability

Whenever possible, the same biological matrix as the intended samples should be used for validation purposes. For some matrices of limited availability, such as bone marrow or blood of a new-born, physiologically appropriate proxy matrices should be used. In this case, the validation is performed in a proxy matrix. Then, cross-validation should be performed as follows: a calibration curve is prepared in

the validated proxy matrix; replicate QC samples, including the LOQ, are prepared in both the proxy matrix and in the matrix to be validated. All QC samples are back-calculated from the same calibration curve. The method is considered cross validated if the precision of QC samples satisfies the acceptance criteria. Moreover, it is important to check the intended matrix for specificity. Special attention should be taken to prepare QC samples throughout tissue assay studies.

However, during analysis of study samples, calibration curves can be prepared in a proxy matrix, but it is recommended that QC samples be prepared in the original matrix.

4. Endogenous drugs

Concerning the assay of endogenous compounds, two cases should be considered:

1. the endogenous compound is exogenously administered as a drug,
2. the endogenous compound concentrations are followed either as pharmacological tracer or in order to evaluate the pharmacokinetic (PK)/pharmacodynamic (PD) relationships.

(1) The difficulty of assaying an administered drug when it is endogenous is complicated by the amounts that are endogenously present. Meanwhile, the quality criteria and the assay validation have to be the same as those mentioned above. However, due to endogenous levels of the compound in the biological matrix, it is difficult to determine the LOQ and the accuracy of the method used. In order to perform the standard curves and the QC sample pools, some solutions are possible:

1. modification of the biological matrix (for example, filtration on activated charcoal–dextran, dialysis, etc.) when possible
2. use of a biological matrix without the endogenous compound to determine (for example, biological parameters varying with circadian rhythm, as for melatonine hormone)
3. use of a substitute matrix (i.e. plasma of another species, human serum albumin, buffers, 0.9% sodium chloride, etc.).

In the last case, the validation procedure should be performed in the substitute matrix, including the LOQ. It is recommended to test the precision (repeatability and reproducibility) on QC samples prepared in the biological matrix under study. It is possible to determine the accuracy throughout the validation step, using the same blank biological matrix to prepare the standard curves, and all pools of QC samples (three concentrations) to be analysed. On the contrary, throughout the pharmacokinetic study, QC samples should be prepared in the biological matrix being investigated and validation of the assay should be carried out using the mean concentration found from six or more assays of each QC sample pool as the reference value or by determining the concentration as follows: $C_{real} = C_{found} - C_{basal}$; where C_{basal} is computed from a calibration curve performed in the substitute matrix. However, in this case, it would be more difficult to have the same criteria for accuracy (i.e., ±20%) and larger criteria should be defined during the validation step.

(2) The quality criteria should be defined with respect to the aim of the study.

1. In order to evaluate a PK/PD relationship, the assays have to be performed according to the same guidelines as for kinetic studies of exogenous compounds.
2. For pharmacological studies, wider quality criteria should be used but they have to be well defined in the experimental protocol.

5. Revalidation and cross-validation

When chromatographic conditions, or sample pretreatment, are modified, revalidation may be necessary. The decision regarding which parameters require revalidation should be based on logical consideration of the specific validation parameters likely to be affected by the change. For example, changes of extraction solvent, buffer or back-extraction matrix

may affect linearity, recovery, selectivity, LOQ, precision and accuracy. A change of the analytical column or mobile phase may affect linearity and selectivity without affecting recovery. Guidelines to revalidation have been proposed by Dadgar et al. [15].

A cross-validation can be carried out by applying a validated method in a given biological matrix to the same type of matrix from another species, or to a different matrix (e.g., plasma and serum) from the same species, or to the same matrix with a change in anticoagulant.

6. Application to drug analysis in pharmacokinetic study

A standard curve including blank matrix should be generated for each analytical run and should be used to determine the sample concentrations in the unknown samples. The same analyst should prepare the calibration curve and the samples to be analysed. Estimation of unknowns by extrapolation of standard curves below the low standard or above the high standard is not recommended. Pools of QC samples (three concentration levels) are prepared at the beginning of the study by an operator that is not in charge of the assay, using working solutions that are different from those used for preparing standard curves and that are stored under the same conditions as the clinical samples to be analysed. For each run, six QCs should be included and analysed. QC samples should be randomly located amongst the clinical samples within a run. They are used to accept or reject the run. According to the published recommendations of the Washington conference, specifically in the field of bioanalytical method validation, at least four of the six QC samples must be within ±20% of their respective nominal values; two of the six QC samples (not both at the same concentration) may be outside the ±20% respective nominal value. When additional QC samples are required for the study, they should be prepared before the original QC sample sets are depleted. Both the original and freshly prepared QC samples should be analysed together, to determine if they are statistically equivalent. Moreover, it is good practice to include blank matrix from the subject, for whom unknown samples are analysed (usually considered QC0), in a run, in order to check the selectivity of the method.

The question of how many calibration curves to run with each sequence of analysis should be answered. If study samples are run singly, a single calibration curve should normally be performed. If replicates of study samples are analysed, identical replication of standard curves is recommended. However, in practice, calibration curves are performed in duplicate even if the samples are analysed singly.

The difficulty arises during biological sample analyses when less than the validated volume is available, and a partial volume used for analysis or when samples analysed are found to be above the calibration curve range, thus requiring dilution for reanalysis. In these cases, both the partial sample volume necessary to bring up the validated volume and the dilution of samples should be performed by addition of a blank matrix. To validate the use of partial sample volume, QC samples are prepared both above the calibration range to demonstrate accurate dilution to within the range, and at a concentration such as the one diluted, the samples are near, but not below the LOQ of the method.

Occasionally, in order to obtain a reliable concentration (>LOQ), it may be desirable to increase the volume of the sample matrix. QC samples can be prepared under the same conditions to validate this procedure. Validation is necessary to show that selectivity is not compromised, and the pre-defined criteria for accuracy and precision are met with the larger sample volume. In addition, selectivity should be verified using increased blank volumes.

There is general agreement that QC samples should be analysed together with the samples. It is recommended that the QC samples are dispersed evenly in a low-high, high-low sequence throughout the batch, in order to detect analytical problems. Criteria of acceptance of QC samples is based on a combined accuracy and precision criterion with an arbitrary range around the "nominal value".

It is debatable whether a sample with a concentration that is incongruous with the pharmacokinetic profile should be reanalysed. In the case of insufficient sample volume, the result should not be reported. However, if a decision is made to repeat

the analysis, it is desirable that the reanalysis be conducted in duplicate. Reporting a single result from multiple determinations requires pre-defined criteria for data selection. On the basis of these criteria, the decision is made to report either the original sample value, or a choice between, for example, the mean of duplicate reanalyses and the mean value of the original analysis and reanalysis. A reanalysis flow-chart has been proposed by Lang and Bolton [10].

7. Conclusion

The essential performance characteristics for bioanalytical validation procedure have been discussed in detail in this paper to provide guidance to bioanalytical chemists. Models used for analytical calibration curves are explained in terms of validity and limitation, along with a presentation of the most common statistical considerations used to validate the model. It is essential to link the impact of validation requirements and acceptance criteria on the determination of bioequivalence and other pharmacokinetic evaluation.

Acknowledgments

The members of the BPL analytical workshops of the "Groupe Métabolisme Pharmacocinétique" are gratefully acknowledged for their scientific assistance.

References

[1] Food and Drug Administration, Draft Guideline for the Format and Content of the Human Pharmacokinetics and Bioavailability Section of an Application. Docket No. 85 D-0275 1985.

[2] E. Purich, in K.S. Albert (Editor), Bioavailability Bioequivalency Regulations: A FDA Perspective in Drug Absorption and Disposition, American Pharmaceutical Association, Washington, DC, 1980.

[3] D. Dradgar and M.R. Smith, Trends Anal. Chem., 5 (1986) 115.

[4] V.P. Shah, Clin. Res. Pract. Drug. Reg. Affairs, 5 (1987) 51.

[5] A.R. Buick, M.V. Doig, S.C. Jeal, G.S. Land and R.D. McDowall, J. Pharm. Biomed. Anal., 8 (1990) 629.

[6] A.G. Causey, H.M. Hill and L.J. Philips, J. Pharm. Biomed. Anal., 8 (1990) 625.

[7] G.P. Carr and J.C. Wahlich, J. Pharm. Biomed. Anal., 8 (1990) 613.

[8] H.T. Karnes, G. Shiu and V.P. Shah, Pharm. Res., 8 (1991) 421.

[9] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Jacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman and S. Spector, J. Pharm. Biomed. Anal., 16 (1991) 249

[10] J.R. Lang and S. Bolton, J. Pharm. Biomed. Anal., 9 (1991) 357.

[11] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Jacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman and S. Spector, J. Pharm. Sci., 81 (1992) 309.

[12] J.W. Hooper, D. Lessard, K. Selinger and D. Dadgar, A New, More Reliable Method for the Determination of the Long Term Frozen Stability of Drugs in Biological Fluids, 8th International AAPS Conference, Orlando, FL, 1993.

[13] International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use, in Draft Consensus Text of Validation of Analytical Procedures, Step 2 of the ICH Process, 1993.

[14] C. Hartmann, D.L. Massat and R.D. McDowall, J. Pharm. Biomed. Anal., 11 (1994) 1337.

[15] D. Dadgar, P. Burnett, M.G. Choc, K. Gallicano and J.W. Hooper, J. Pharm. Biomed. Anal., 13 (1995) 89.

[16] International Union of Pure and Applied Chemistry, Anal. Chem., 48 (1976) 2294.