ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
14 (1995) 23–31

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Issues in evaluation of bioanalytical method selectivity and drug stability ☆

Darioush Dadgar *, Pamela E. Burnett

Analytical Department, Phoenix International Life Sciences, 4625 Dobrin Street, St. Laurent (Montreal), Quebec, H4R 2P7, Canada

Received for review 25 April 1995; revised manuscript received 31 July, 1995

Abstract

Evaluation of selectivity is one of the most controversial aspects of method validation and application of methods to studies. The focus of selectivity testing should address the question: Above what level will interference significantly impact on study conclusions? Four key issues will be addressed: the statistical relevance of any selectivity test performed; a criterion for significant interference; experimental methods to establish selectivity; and criteria for acceptance.

To ensure that compound integrity is maintained throughout the work-up process, statistically meaningful methods of stability evaluation which are associated with specific acceptance criteria are required. Suitable methods for evaluating stability of analyte and/or solutions of analyte, in process stability, processed sample stability, long term stability and freeze–thaw stability, as well as meaningful acceptance criteria, are presented.

Keywords: Selectivity; Stability; Chromatography; Bioanalysis; Confidence intervals; Statistical analysis

1. Introduction

The validation of bioanalytical methods has been the subject of wide discussion in recent papers [1–7]. The most widely observed guiding principles for validation have been based on a 1992 conference report [1]. While this report provides guidelines on the parameters requiring validation in bioanalysis, it does not provide specifics on how validation can be accomplished. This paper deals with two issues in validation. The first concerns the evaluation of method selectivity and interferences with the statistical use of data obtained for blank matrix samples to set the limit of quantitation (LOQ). The second discusses a practical approach to evaluation of analyte stability, and the statistical treatment of data obtained. While the concepts presented in this paper may be applicable to a wide variety of bioanalytical tech-

☆ Presented at the Sixth International Symposium on Pharmaceutical and Biomedical Analysis, April 1995, St. Louis, MO, USA.

* Corresponding author.

SSDI 0731-7085(95)01634-1

niques, they are primarily intended for application to chromatographic techniques including HPLC, GC, LC/MS, GC/MS, and LC/MS/MS.

2. Evaluation of method selectivity

An important emphasis in method validation is the demonstration of method specificity. A method is said to be specific if it produces a response for only a single analyte. Method "selectivity" is the ability of a method to produce a response for the target analyte distinguishable from all other responses. Since most chromatographic methods produce responses not only for the analyte(s) of interest, but also for other substances, the term "selectivity" is usually more appropriate in this context than "specificity". Therefore, the term "selectivity" is used here [2].

Interference with the assay of analytes in biological samples typically arises from a number of endogenous and exogenous sources, including the following.

Endogenous: analyte metabolites and/or precursors; analyte degradation products; co-administered drugs, vitamins and/or their metabolites and/or degradation products; chemicals normally occurring in the biological fluid (e.g. hormones, proteins, lipids, dietary substances, etc.).

Exogenous: impurities in reagents used for sample work-up; substances used in the manufacture of labware (e.g. plasticizers), or resulting from incomplete washing of labware.

2.1. The statistical relevance of interference

While ideally zero percent interference with analytes of interest is desired this may not always be achievable, so that in practice minor interferences may be allowable. Interference levels must be evaluated relative to the levels of analyte to be measured. Before beginning method development, the following factors should be defined, at least tentatively: the lowest concentration of analyte which must be accurately measurable, i.e. the "desired" LOQ; the level at which interference with the assay significantly alters measured values, especially for low concentration samples, with implications for study conclusions. In addition, target criteria for accuracy and reproducibility should be set.

Simple selectivity tests are routinely conducted during method development to help guide the researcher toward a "promising" analytical method. However, once a method is at the validation stage, the selectivity test should assess level and reproducibility of interference, while simultaneously assessing the impact of interference on precision and accuracy of the method at the LOQ. The response of the LOQ standard should be greater than that for the response from the blank biological matrix by a defined factor, as discussed below.

2.2. Definition of unacceptable interference

Interference is considered "unacceptable" when it affects the accuracy of quantitation at concentrations near the LOQ, to an extent that may result in erroneous study conclusions.

When precision and accuracy criteria at the LOQ are not met, options are limited. (i) Adjust the LOQ to take into account the higher than expected interference. This approach compromises method sensitivity in order to achieve the desired selectivity. This option may be considered if, for example, the LOQ has been targeted substantially lower than the expected concentrations of real samples, or in evaluation of pharmacokinetic profiles if the LOQ has been set more than the minimum desired number of elimination half lives below the observed C_{max} value. (ii) Continue method development to improve method sensitivity, selectivity and/or reproducibility. If chromatographic columns and conditions as well as sample clean-up techniques have been investigated thoroughly and failed to improve method selectivity, switching to a more selective detector such as MS, or MS/MS may prove successful. (iii) Revise the acceptance criteria. This latter option is to be avoided unless criteria have been more tightly set than the generally accepted criteria for the analytical technique in question. An example of such criteria is outlined in the 1992 conference report [1] (accuracy 80–120%, RSD $\leqslant$ 20% at the LOQ for chromatographic techniques).

2.3. Experimental methods to evaluate selectivity

The following practical approaches may be used during method development, to investigate the selectivity of an analytical method.

(i) Processing blank matrix samples from independent sources is usually sufficient to demonstrate lack of potential interference from substances native to the biological fluid. However, this will not provide evidence for lack of interference from drug metabolite(s).

(ii) If known metabolites and degradation products can be obtained commercially or synthesized, they can be added to blank biological matrix and any potential interference assessed. In the absence of such metabolites or degradation products, a pilot study can be conducted in which subjects or animals are dosed and biological samples analyzed by chromatography under high resolution ("stretched") or gradient conditions, to look for potential metabolites partially resolved from under the analyte peak(s). If analyte concentrations are sufficiently high, and the chromophores differ sufficiently, the use of diode array or scanning UV detection under the regular (not "stretched") conditions can give evidence for peak purity.

(iii) Potential metabolites can be produced in vitro by incubation with liver homogenates, and chromatographed to check for potential interference with analyte(s) of interest.

(iv) To assess the effect of potential acid hydrolysis products in the stomach, in vitro incubation of the analyte in acidic medium can be conducted followed by neutralization and chromatography.

(v) Processing of "reagent blanks" in the absence of biological matrix (or with an equivalent volume of pure water) is normally adequate to demonstrate selectivity with regard to the exogenous interferences mentioned above, using the same labware, reagents, volumes, etc. If interference is observed, a series of experiments is designed to identify the source.

Once any interferences from analyte metabolites and degradation products have been defined and minimized during method development, preliminary evaluation of the blank biological matrix follows. A study should be designed to provide a sound statistical evaluation of the extent of interference from biological matrix originating from different sources. In practice, interference in predose study samples is one of the most commonly observed problems with otherwise validated bioanalytical methods once they are put into routine use. This arises from evaluation of an insufficient number of sources of blank matrix during validation. While the Washington conference report recommends evaluation of a minimum of six matrix sources [1], in practice evaluation of 10–20 sources of blank matrix may be necessary to allow meaningful confidence intervals about the mean interference level to be constructed. This also improves the likelihood of discovering every selectivity problem during validation, as opposed to during analysis of real samples.

The following study design may be used to evaluate selectivity with regard to different sources of biological matrix.

(i) Prepare a calibration curve in a biological matrix previously shown to be free of detectable interferences with the analyte(s) of interest. Standard concentrations should include an LOQ standard and an appropriate number of higher concentration standards covering the range to be validated.

(ii) Obtain the biological matrix from at least 10 independent sources.

(iii) Spike each blank matrix individually with analyte(s) at a concentration equivalent to the proposed LOQ standard(s) (LOQ QCs).

(iv) Process the calibration curve, the blank biological matrix samples and the spiked LOQ QCs, adding internal standard if applicable.

(v) Determine the apparent responses for the blank biological samples, and the responses for the calibration curve and the proposed LOQ QCs. Calculate the concentrations for the LOQ QC standards.

2.4. Acceptance criteria for interference

Many laboratories set acceptance criteria for interference in terms such that the response for a blank matrix sample must not exceed a predefined percentage of that for the LOQ standard. This percentage is often set arbitrarily (e.g. 20%,

25% or 50%), without real consideration of the potential impact on the determination of analyte concentrations near the LOQ.

Kucharczyk [3] suggests that "The lower limit of quantitation is that concentration of the analyte in the matrix of interest for which the confidence interval at the 95% (or any other) probability level does not overlap with the confidence interval of the matrix blank standard.". He extrapolates the predicted confidence intervals around the regression line to determine the confidence intervals at the zero (blank) standard and at any analyte concentration X. The value of X for which the confidence interval does not overlap with that of the blank standard is defined by him as the LOQ [3].

Although this approach has merit, two weaknesses are apparent.

(i) Since levels of interference typically vary from one source of biological matrix to another, it is not sufficient to predict the LOQ based on replicate calibration curves prepared in a single blank matrix source or pool, as Kucharczyk has apparently done.

(ii) The mere fact that the confidence intervals of the LOQ and the blank standard do not overlap does not necessarily provide sufficient evidence for selectivity in quantitation of unknown samples at the LOQ. Intrinsic to this approach is a 2.5% probability that the response of the zero (blank) concentration standard will be the same as or greater than that of the LOQ standard.

This approach may be strengthened by the following proposal. Construct 95% confidence intervals around the responses for the multiple zero (blank) standards, and for the replicate LOQ standards. The signal-to-noise ratio (analyte response/average noise of an adjacent area of the chromatogram showing no peak) of an LOQ under consideration should be better than a predefined factor, e.g. three, five, or better.

For the matrix blank standards giving a mean response of X units, or a proposed LOQ standard (with a mean response of X units), the upper (UL) and lower (LL) limits of the 95% confidence bands are expressed as follows:

$$\mathrm{UL} = X + t_{0.975}(s/n^{1/2})$$

$$\mathrm{LL} = X - t_{0.975}(s/n^{1/2})$$

where $t_{0.975}$ = the value of the Student t distribution with $n-1$ degrees of freedom, s = the standard deviation, and n = the number of measured values.

It is proposed that the LOQ should be defined as the analyte concentration for which the lower confidence limit of the mean response is equal to or greater than four times the upper confidence limit for the mean blank standard response. A factor greater than four could be selected. The average interpolated concentration of the proposed LOQ should lie between 80% and 120% of its theoretical concentration and the RSD should be $\leqslant 20\%$.

While it would be preferable to construct full calibration curves for each source of biological matrix (use of predicted confidence interval bands around the entire curve accounts for variation at the LOQ caused by imperfect fit of standards) this would put an impractical cost/time burden onto the validation process. For this reason, it is suggested that the evaluation of the mean responses of blank standard and of the proposed LOQ standard should be adequate for the purpose of defining an appropriate LOQ.

Two examples based on hypothetical data are presented in Tables 1 and 2. Each illustrates the different situations that can arise in practice.

In Table 1 LOQ # 2 corresponding to twice the response (and therefore concentration) of LOQ # 1 is required to fulfil the acceptance criterion (factor 4).

In Table 2, where the blank matrix response is more consistent (RSD is ten times less than in Table 1) and the assay RSD is low, the LOQ required to fulfil the acceptance criterion is ≈ 1.3 times more concentrated than LOQ # 1. In both cases LOQ #1 produces comparable mean responses.

3. Stability evaluation

An essential aspect of method validation is to demonstrate that analyte(s) is (are) stable in the biological matrix and in all solvents encountered

Table 1
Calculation of 95% confidence intervals for hypothetical matrix blank standard responses and two proposed LOQ standard responses: matrix blank interference inconsistent, assay RSD values high

Biological matrix source	Matrix blank standard response	Proposed LOQ #1 response	Matrix blank std. response (% of LOQ #1)	Proposed LOQ #2 response	Matrix blank std. response (% of LOQ #2)
#1	126	176	71.6	242	52.1
#2	0	130	0.0	347	0.0
#3	0	125	0.0	304	0.0
#4	30	175	17.1	268	11.2
#5	87	160	54.4	252	34.5
#6	0	158	0.0	327	0.0
#7	0	147	0.0	287	0.0
#8	50	155	32.3	244	20.5
#9	33	135	24.4	356	9.3
#10	0	100	0.0	249	0.0
Mean	32.6	146.1	20.0	287.6	12.8
SD	43.85	23.80		43.65	
RSD (%)	134.5	16.3		15.2	
Confidence limits					
Upper 95%	63.9	163.1		318.8	
Lower 95%	1.3	129.1		256.4	
LOQ Std. lower conf. limit > 4 × matrix blank std. upper conf. limit?		No		Yes	

during the sample work-up process, under the conditions to which study samples will be subjected. Stability studies are best conducted at a minimum of two analyte concentrations (high and low).

Ideally samples stored for stability evaluation should be obtained from dosed subjects. If, however, metabolite reversion to drug, or ongoing metabolism in the frozen matrix are not problems, blank matrix samples spiked with analyte are an acceptable and more commonly used alternative.

3.1. Types of stability evaluation required

A number of areas can be identified where stability evaluation is required.

(i) Long term stability of analyte(s) in the biological matrix should be evaluated:

- concurrently with the stability of a matrix blank, to determine whether matrix degradation gives rise to interferences;
- in a way that demonstrates that samples are stable over the maximum period of study sample storage;
- at the temperature of study storage;
- in the container type in which study samples will be stored;
- separately for each applicable matrix; analyte stability cannot be extrapolated between species for the same matrix (e.g. a demonstration of stability in rat plasma cannot be considered to indicate stability in mouse plasma).

(ii) Stability of reference standard stock solutions and their dilutions should be evaluated over the maximum period for which reference standard stocks and dilutions will be stored prior to use, at the same temperature at which stocks to be used during the study will be stored, and in the same solvent and container type used for stocks to be used during the study.

(iii) Short term stability of analyte(s) in the biological matrix should be evaluated following

Table 2
Calculation of 95% confidence intervals for hypothetical matrix blank standard responses and two proposed LOQ standard responses: matrix blank interference consistent, assay RSDs low

Biological matrix sources	Matrix blank standard response	Proposed LOQ #1 response	Matrix blank std. responses (% of LOQ #1)	Proposed LOQ #2 response	Matrix blank std. response (% of LOQ #2)
#1	38	157	24.2	205	18.5
#2	40	121	33.1	197	20.3
#3	35	143	24.5	189	18.5
#4	46	140	32.9	168	27.4
#5	36	159	22.6	198	18.2
#6	37	137	27.0	210	17.6
#7	35	124	28.2	185	18.9
#8	37	142	26.1	196	18.9
#9	47	141	33.3	165	28.5
#10	49	151	32.5	201	24.4
Mean	40.0	141.5	28.4	191.4	21.1
SD	5.31	12.42		14.95	
RSD (%)	13.3	8.8		7.8	
Confidence limits					
Upper 95%	43.8	150.4		202.1	
Lower 95%	36.2	132.6		180.7	
LOQ std. lower conf. limit $>4 \times$ matrix blank upper conf. limit?		No		Yes	

storage under laboratory conditions used for sample work-up (e.g. for stable compounds, at room temperature; for unstable compounds, on ice or stabilized by other means) for a period of e.g. 6 h or 24 h, and compared with data from the same samples prepared (or thawed) and analyzed without delay.

(iv) In-process stability. For compounds found to be unstable during the short-term stability evaluation, separate evaluations of analyte stability in each matrix encountered during sample processing (e.g. buffer, extraction solvent, back-extraction solvent) are useful to help pinpoint the cause(s) of the instability. Measures can then be taken to attempt to stabilize the analyte during "problem" steps.

(v) Processed sample stability should be evaluated over the maximum time from completion of sample work-up to completion of data collection (e.g. for chromatographic methods the time from completion of sample extraction to completion of chromatography), with an allowance for potential delay in analysis due to equipment failure. This stability study is conducted at the temperature at which processed study samples will be held prior to data collection (e.g. refrigerated or room temperature corresponding to the autosampler temperature).

(vi) Freeze–thaw stability of analyte(s) in the biological matrix should be evaluated in one of two ways: (a) after the matrix is frozen for a specific period and thawed three times (three freeze–thaw cycles) and compared with data obtained from fresh unfrozen samples; or (b) after each of three freeze–thaw cycles and compared with data obtained from fresh unfrozen samples. If approach (a) is used and the analyte is found to be unstable, then the matrix should be analyzed

after each freeze–thaw cycle, in order to establish after how many cycles it is adequately stable. If, for example the analyte is found to be stable after only one cycle of freeze–thaw, then measures must be taken during study sample collection to store an adequate number of sample aliquots to permit repeats, without having to freeze and thaw the sample more than once.

Potential measures to stabilize analytes in biological matrices

A number of measures are available to reduce the loss of analyte during storage.

(i) Freezing at lower temperature (e.g. −80°C, or in liquid nitrogen) to slow down the rate of degradation, as suggested earlier by Buick et al. [6].

(ii) Use of appropriate additives (e.g. antioxidants, enzyme inhibitors, buffers) to stabilize analyte in the biological matrix.

(iii) Immediate derivatization following sample collection to form a stable derivative (e.g. captopril in whole blood).

(iv) Immediate extraction of analyte from the biological matrix for storage either dry or in a solvent in which analyte has been proven to be stable. If all other approaches fail:

(v) Immediate extraction followed by prompt analysis of samples.

3.2. Procedures for evaluating stability

Typically laboratories evaluate analyte stability in a given matrix in one of the following ways, usually processing stability samples in duplicate or triplicate. (i) Samples of the blank biological matrix are spiked with analyte at a given level and stored for a specified period of time. After the storage period, a second set of blank samples is spiked at a similar level using fresh stocks, and immediately analyzed together with the stored samples, using a freshly prepared standard curve. Interpolated values for the fresh and stored samples are then compared directly. (ii) Blank samples spiked with analyte at a specific level are stored for the desired period of time, and evaluated against a freshly prepared standard curve. The analyte is considered stable if a predetermined number of stored QCs meet the specified QC acceptance criteria. (iii) Biological matrix spiked with analyte is analyzed immediately after preparation, and aliquots are stored for one or more periods at a specified temperature, before analysis using a freshly prepared standard curve on each occasion. Concentrations determined on each occasion are compared to detect any trend in degradation.

Confidence intervals are wide and unreliable when only two or three replicates are processed, making the probability of erroneous conclusions from the data high. In addition, each of the above methods suffers from errors, which can be categorized as processing variability error, and bias error. Processing variability errors are due to variations inherent in the analytical method (pipetting, extraction, transfer losses, chromatography). Processing variability error is reflected in within-run RSD data and between-run RSD data. Approaches (i) and (ii) above suffer from within-run variation. Approach (iii) suffers from within-run and between-run variation. Bias error arises during the preparation of calibration standards and QC samples and is reflected in the systematic deviation of the actual concentrations prepared from the nominal (intended) concentrations. Bias error is involved in all the above methods, either through repeated spiking of calibration curves and/or QC samples, but it can be minimized.

The authors propose the following approach to stability evaluation: a modification of method (i) above, based on an approach published by Timm et al. [4]. This can minimize many of the errors noted above, while providing statistically meaningful data on which to draw conclusions. This approach can be applied to evaluation of each of the criteria for stability discussed previously. The following example describes the evaluation of long-term stability of analyte in the biological matrix under specified freezing conditions. Blank biological matrix is spiked with analyte at two concentrations (high and low), and stored under the required conditions. These are termed the "stability samples". Following the designated storage period, fresh biological matrix is spiked with analyte at the same two concentrations as the stability samples; these are termed the "com-

parison samples". Ten replicates of each of the two stability samples and the two comparison samples are simultaneously analyzed as a single batch, and the 40 responses are determined.

Advantages. Fresh and stored samples are analyzed simultaneously, eliminating between-run variation. No calibration curve is used thus avoiding any bias associated with the use of separate curve preparations. Processing 10 replicates provides adequate data to construct meaningful confidence intervals about the mean responses for each of the four sets of samples analyzed.

Disadvantage. This approach suffers from bias error associated with separate preparations of fresh and stored samples (this applies to long-term stability and stock stability only). This disadvantage, however, is currently unavoidable since regulators have insisted on the need for processing either freshly prepared calibration standards or freshly prepared QC samples together with stability samples.

An alternative strategy proposed by the authors involves the preparation of a single batch of fresh biological matrix spiked at an appropriate level with analyte and then subdivided for use both as a "stability" and a "comparison" sample. A portion of the spiked matrix is stored under study sample storage conditions (the "stability" sample), while the other portion is stored at −130°C, to serve as the "comparison" sample(s), proposed as being equivalent to freshly prepared samples. Studies conducted at Phoenix International Life sciences, as yet unpublished, have shown that even highly unstable drugs such as acetylsalicylic acid are stable in certain biological matrices for months at −130°C [8]. However, the possibility of matrix effects on stability even at this low temperature should be considered.

3.3. Acceptance criteria

In this proposal, the analyte may be considered adequately stable if:

(i) the 90% confidence limits for the mean response of the combined low and high concentration stability samples about the mean for the comparison samples meet the following criteria:

- the lower confidence limit is $>85.0\%$
- the upper confidence limit is $<115.0\%$

(ii) the ratios of the mean response for the stability samples to that of the comparison samples, for:

- the combined high and low QCs
- the low QC, and
- the high QC

all lie within the range 90–110%.

Acceptance limits other than those suggested here may be appropriate; however, as with all other validation criteria, it is essential that the desired limits be set a priori. The authors are currently assessing the utility of this proposal, and the necessary statistical package, for future publication.

4. Conclusions

Discussion on issues pertaining to the validation of bioanalytical methods and acceptance criteria has been evolving in recent years. To date, acceptance criteria applied to bioanalytical data have been largely based on "historical" arbitrary approaches, and have resulted in relatively minor variations in criteria between laboratories. In this paper, issues related to bioanalytical method selectivity and drug stability have been discussed and rational acceptance criteria have been defined. It is hoped that this publication will prompt yet further discussion on these and other related validation topics.

Note

The procedures proposed and outlined in this paper are not necessarily in practice at the institution with which the authors are affiliated.

Acknowledgement

The authors wish to acknowledge Professor Anthony Fell for his editorial contribution to this paper, and Deborah Frankland for secretarial assistance.

References

[1] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathian, C.E. Cook, R.D. McDowall, K.A. Pittman and S. Spector, J. Pharm. Sci., 81 (1992) 309–312.
[2] H.T. Karnes, G. Shiu and V.P. Shah, J. Pharm. Res., 8 (1991) 421–426.
[3] N. Kucharczyk, J. Chromatogr., 612 (1993) 71–76.
[4] U. Timm, M. Wall and D. Dell, J. Pharm. Sci., 74 (1985) 972–977.
[5] A.G. Causey, H.M. Hill and L.J. Phillips, J. Pharm. Biomed. Anal., 8 (1990) 625–628.
[6] A.R. Buick, M.V. Doig, S.C. Jeal, G.S. Land and R.D. McDowall, J. Pharm. Biomed. Anal., 8 (1990) 629–637.
[7] D. Dadgar, P.E. Burnett, M.G. Choe, K. Gallicano and J.W. Hooper, J. Pharm. Biomed. Anal., 13 (1995) 89–97.
[8] J.W. Hooper, D. Lessard, K. Selinger, R. Parkes and D. Dadgar, A new method for determining the stability of drugs in biological fluids, and a confidence interval approach to statistical evaluation of the data, poster presentation, 8th Annual AAPS Meeting, 1993.