ELSEVIER

Journal of Chromatography A, 730 (1996) 381–394

JOURNAL OF CHROMATOGRAPHY A

Rational experimental design for bioanalytical methods validation
Illustration using an assay method for total captopril in plasma

J. Wieling*, G. Hendriks, W.J. Tamminga, J. Hempenius, C.K. Mensink, B. Oosterhuis, J.H.G. Jonkman

Pharma Bio-Research International BV, Science Park, NL-9471 GP Zuidlaren, Netherlands

Abstract

Generally, bioanalytical chromatographic methods are validated according to a predefined programme and distinguish a pre-validation phase, a main validation phase and a follow-up validation phase. In this paper, a rational, total performance evaluation programme for chromatographic methods is presented. The design was developed in particular for the pre-validation and main validation phases. The entire experimental design can be performed within six analytical runs. The first run (pre-validation phase) is used to assess the validity of the expected concentration–response relationship (lack of fit, goodness of fit), to assess the specificity of the method and to assess the stability of processed samples in the autosampler for 30 h (benchtop stability). The latter experiment is performed to justify overnight analyses. Following approval of the method after the pre-validation phase, the next five runs (main validation phase) are performed to evaluate method precision and accuracy, recovery, freezing and thawing stability and over-curve control/dilution. The design is nested, i.e., many experimental results are used for the evaluation of several performance characteristics. Analysis of variance (ANOVA) is used for the evaluation of lack of fit and goodness of fit, precision and accuracy, freezing and thawing stability and over-curve control/dilution. Regression analysis is used to evaluate benchtop stability. For over-curve control/dilution, additional to ANOVA, also a paired comparison is applied. As a consequence, the recommended design combines the performance of as few independent validation experiments as possible with modern statistical methods, resulting in optimum use of information. A demonstration of the entire validation programme is given for an HPLC method for the determination of total captopril in human plasma.

Keywords: Validation; Experimental design; Pharmaceutical analysis; Chemometrics; Captopril; Paracetamol

1. Introduction

Bioanalytical methods must be validated if the results are used to support the registration of a new drug or the reformulation of an existing one. The validation is required to demonstrate the performance of the method and the reliability of analytical results. It is essential to use well characterized and fully validated methods to yield reliable results that can be interpreted

* Corresponding author. Present address: Bio-Intermediair Europe BV, P.O. Box 454, NL-9700 AL Groningen, Netherlands.

0021-9673/96/$15.00
PII S0021-9673(96)00006-4

satisfactorily. Many papers have appeared recently on the validation of bioanalytical methods. The most relevant of these are referred to here [1–5]. The conference report by Shah et al. [4] (the Washington Conference Report) is generally accepted as one of the most important guidelines for bioanalytical methods validation. The conference was attended by many workers in the field from pharmaceutical industries, regulatory bodies, contract research organizations and universities and many of them contributed to the report. Nevertheless, some critical notes to certain aspects of the guidelines in the Washington Report have been made by Hartmann et al. [5]. They have extracted some recommendations after analysis of the Washington criteria, namely that it would be preferable for the terminology to be consistent with the existing guidelines in other fields of chemical analysis. More importantly, the authors drew attention to statistical considerations and to the nature of experimental errors and the separation of the total measurement error into its constant (bias) and its random (precision) components.

An excellent paper on the practical issues of bioanalytical methods validation was published by Dagdar et al. [6]. They discussed all aspects of bioanalytical methods validation, from prerequisites to method validation to revalidation and cross-validation. Useful procedures are described for the long-term stability validation of biological samples. Interesting comments on other stability questions are included and ideas on ruggedness testing and internal standard selection.

For research and development laboratories within a service, the validation of a method should be cost-effective and results should be available as soon as possible. Therefore, the experimental set-up of the validation study should be efficient and sensible. This paper deals with the design and the experimental set-up of the validation of bioanalytical methods within our laboratory. The design was originally applied to chromatographic assays, but may, after a few adaptations, also be used for other analytical techniques, such as immunoassays. Particular attention will be given to statistical analysis of validation results. This paper will not emphasize items such as the preparation of standards for calibration, quality control and validation or strategies for method optimization. Other workers have highlighted these subjects in the context of bioanalytical methods validation [1–3]. Definitions of the required performance characteristics (validation criteria) are also not included; for descriptions we refer to other publications [1–5]. These papers also discuss in detail the background, the meaning and the rationale for the characteristics to be validated.

2. Validation study protocol

All analytical methods developed in our laboratories are validated comprehensively, and all aspects with regard to specificity, sensitivity, calibration model, recovery, accuracy, precision, stability and overcurve control of samples are covered. When bioanalytical studies are performed, the relevant international guidelines, recommendations and requirements are taken into account as comprehensively as possible. For the development and validation of the assay method, this concerns the Note for Guidance on Analytical Validation [7] and the Conference Report on Analytical Methods Validation (Shah et al. [4]). Development and validation of assay methods are conducted in accordance with current good laboratory practice standards [8,9]. Proper interpretation of these guidelines and regulations directs investigators towards the design of protocols for all studies to be performed in support of drug registration. Studies also include methods validation, hence a signed protocol for assay method validation should be in place prior to the conduct of a validation study. In general, the protocol should be followed. Scientifically justified changes can be made, if the changes are documented and authorized by the study director.

Usually, in protocols for assay method validation we include items such as identification of the analyte(s), including sample matrix and concentration range to be applied, the preparation of calibration, quality control and validation samples, the performance characteristics to be evalu-

ated (including the limits to be maintained), the procedures for evaluation (number of replicates, number of batches/runs, statistics), the concentration range to be evaluated, data filing and reporting.

3. Validation programme

3.1. Analytical method

We demonstrate here our validation programme by means of a recently developed method for the determination of total captopril concentrations in human plasma. Captopril is a potent and selective inhibitor of angiotensin-converting enzyme (kininase II). Oral captopril administration lowers blood pressure in hypertensive humans. Captopril is readily converted into its disulphide dimer and forms disulphide conjugates with endogenous thiol compounds [10]. Only the free captopril is pharmacologically active; however, the formation of the inactive disulphides is reversible; subsequently, they may act as a reservoir of free captopril and contribute to a longer duration of action than predicted by the blood concentrations of free captopril [11]. As a consequence, the total captopril plasma concentration is probably an important parameter in relation to therapeutic effect. To measure total captopril (free captopril + captopril disulphides), all captopril disulphide conjugates have to be reduced to free captopril. Immediately after this reduction, a chemical stabilizer must be added to the biological samples, to prevent the re-formation of disulphides. In the method for the quantitative determination of total captopril in human plasma presented in this report, validation samples were used, which had been prepared by spiking blank human plasma with known amounts of captopril–captopril disulphide. No internal standard was used. The requirements for the method had been documented in a Study Protocol for Assay Method [12].

Captopril disulphides in plasma were reduced to free captopril with tris(2-carboxyethyl)phosphine (TCEP). The free captopril was captured with N-pyrenylmaleimide (NPM), in order to protect the free thiol group from the re-formation of disulphides. After the TCEP reduction and after NPM treatment, the reaction mixture was washed to remove interferences. Further sample clean-up was performed by means of liquid–liquid extraction of the NPM adduct of captopril. Separation was performed by reversed-phase high-performance liquid chromatography and quantification with fluorescence detection. The detection limit of the method was approximately 5 ng ml^{-1} using 500 μl of plasma. The lower and upper limits of quantification were 10.0 and 1005 ng ml^{-1}, respectively.

3.2. Validation experimental design

In principle, a common validation study within our laboratories consists of six analytical runs (during a period of at least 72 h), optionally extended by additional analytical runs for the validation of less common performance characteristics. Here, an analytical run is defined as a batch of samples consisting of a test sample for system suitability, calibration samples ($\geq$8), quality control samples ($\geq$6), a blank sample and clinical study samples ($\geq$0), which can be analysed within a predefined period of time (commonly 24 h).

A diagram of the entire procedure for the bioanalytical methods validation is given in Fig. 1. For each validation study, the pre-validation phase (run 0) is used to assess the validity of the expected concentration–response relationship (lack of fit, goodness of fit), to assess the specificity of the method and to assess the benchtop stability of processed samples. The latter experiment is performed to justify overnight analyses.

After approval of the suitability of the method after the pre-validation phase, the main validation phase (the next five runs) is performed to evaluate method precision and accuracy, recovery, freezing/thawing stability and over-curve control/dilution. An efficient and rational design has been developed for this main validation phase. The design is nested, i.e., many experimental results are used for the evaluation of several performance characteristics. Below, de-

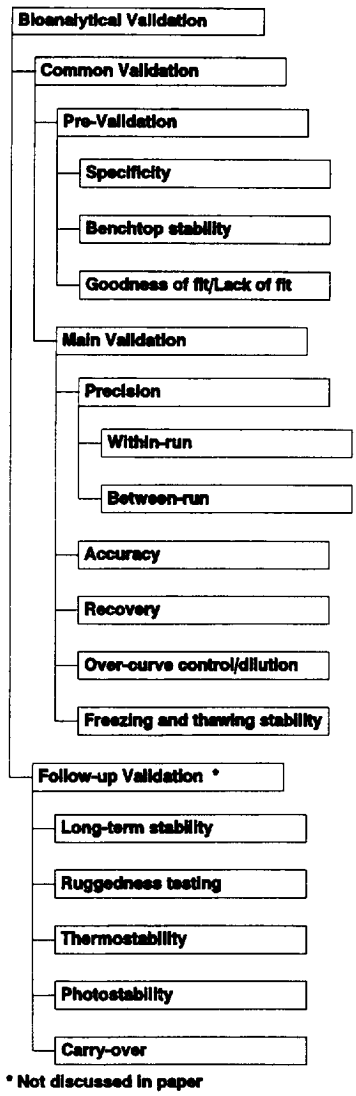

Fig. 1. Diagram of the entire procedure for bioanalytical methods validation.

tails of the entire programme are given. The complete validation study is presented in a scheme in Table 1. If applicable and relevant, follow-up validation runs may be carried out. Here, follow-up validation will not be discussed.

Pre-validation

In run 0 the following samples were analysed:

(i) 24 calibration samples; eight concentrations (10.0, 25.1, 50.2, 100, 251, 502, 754, 1005 ng ml^{-1}) distributed over the concentration range with increasing intervals and analysed in triplicate (entire procedure).

(ii) samples for the assessment of the stability during storage in the sample compartment (benchtop stability); pooled extracts of spiked plasma samples at two concentrations (approximately 25.0 and 750 ng ml^{-1}) were injected every 2 h for a total period of 30 h, during which the extracts were kept in the sample compartment of the injector (protected from light and at a temperature of 10°C).

The specificity of the assay method was checked by analysing at least six independent blank plasma samples. The chromatograms of these blank plasma samples were compared with chromatograms obtained by analysing test solutions of the pure compound treated with NPM. The specificity of the method was also checked for other compounds, namely drugs (and their metabolites, if applicable) that were used as co-medications during drug interaction studies. The above-mentioned experiments for pre-validation are general experiments, especially those for specificity testing. Additional or modified experiments may be performed for reasons such as the inavailability of reference materials for metabolites. Suggestions for solving such difficulties have been given by Dagdar et al. [6] in their prerequisites to methods validation.

Main validation

In runs 1–5 the following samples were analysed:

(i) eight calibration samples containing eight concentrations (10.0, 25.1, 50.2, 100, 251, 502, 754, 1005 ng ml^{-1}) distributed over the concentration range with increasing intervals;

(ii) twelve precision and accuracy samples: four concentrations (10.0, 25.1, 251 and 754 ng ml^{-1}) in triplicate.

(iii) three over-curve control samples; one concentration (2008 ng ml^{-1}) diluted five times in triplicate;

(iv) six samples for stability assessment after repeated freezing and thawing; two concentrations (25.1 and 754 ng ml^{-1}) in triplicate [sub-

Table 1
Scheme of the validation study: number of assays per analytical run for validation

Run	Linearity	Specificity	Stability (30 h)
0	24 calibration samples (8 concentration levels in triplicate)	6 independent blank plasma samples; test samples; co-medication samples	32 injections of pooled sample extracts (every 2 h, at two concentrations)
Total	62 injections		

Run	Calibration samples (8 calibration levels)	Validation samples										Total
		LLQ	X_1			X_2			X_3			
		P&A[a]	P&A[a]	F/T	Rec[b]	P&A[a]	Rec[b]	Dil	P&A[a]	F/T	Rec[b]	
1	(8)	3	3	3[c]	3[c]	3	3[c]	3[c]	3	3[c]	3[c]	38
2	(8)	3	3	3[c]	3[c]	3	3[c]	3[c]	3	3[c]	3[c]	38
3	(8)	3	3	3[c]	3[c]	3	3[c]	3[c]	3	3[c]	3[c]	38
4	(8)	3	3	3[c]	3[c]	3	3[c]	3[c]	3	3[c]	3[c]	38
5	(8)	3	3	3[c]	3[c]	3	3[c]	3[c]	3	3[c]	3[c]	38
Total	190 injections											

[a] P&A = precision and accuracy determinations.
[b] Experiments are direct injections with concentrations equal to 100% recovery in extracts.
[c] Results are used to calculate the performance characteristic using the precision and accuracy data of the same analytical run at the same concentration level.

samples were taken from samples which were prepared from the validation pools and which were frozen and thawed before each next freezing and thawing cycle (run)];

(v) no recovery measurements were performed for total captopril analyses, since a recovery experiment would not only include the liquid–liquid extraction procedure, but also the reduction of disulphides and the derivatization with NPM. Therefore, for simplicity reasons, a demonstration of recovery measurements and data processing with the same validation experimental design is shown for paracetamol (acetaminophen) in plasma here: nine direct injections of test solutions for recovery determination were performed; these direct injections contained such amounts of the analytes as would correspond with 100% recovery from validation samples at three different concentrations (1.00, 20.0 and 35.0 μg ml^{-1}) in triplicate. The recovery of theophylline (internal standard for paracetamol) was evaluated at the concentration used during the actual analysis of the plasma samples (40.0 μg ml^{-1}).

4. Data processing

Calculations for the determination of the validation parameters were performed using spreadsheets programmed in Lotus 123 on IBM-compatible computers. These spreadsheets hold the analysis of variance tables for the determination of precision, accuracy, goodness of fit, lack of fit, freezing/thawing stability and over-curve control/dilution.

The peak height of the captopril–NPM adduct was taken as the response for a given sample. Calibration graphs were calculated by weighted linear regression ($W = X^{-1}$) on the responses of a series of calibration samples versus the corresponding nominal concentrations. The measured concentration in a sample was calculated by substituting the response for that sample in the equation of the corresponding calibration graph.

The calibration data from runs 1–5 were subjected to the following acceptance criteria. A calibration point was rejected as an outlier if the back-calculated concentration for a calibration sample (on the basis of the equation of the

corresponding calibration graph) deviated more than 15% from the nominal value for the two lowest concentrations and more than 10% for the other concentrations. A calibration graph was accepted unless there were more than two outliers, or if there were two outliers on adjacent concentrations.

The validation data from runs 0–5 were subjected to the Grubbs test [13] for the detection of outliers. Outliers, if any, were excluded from the calculation of performance characteristics. The measured concentrations were rounded to three significant digits.

4.1. Pre-validation

Benchtop stability

The measured peak heights for the assessment of the stability in the sample compartment were plotted versus time. The data were used for regression analysis to estimate an increase or a decrease in the measured peak heights. A decrease or an increase of 10% in the measured peak height (based on regression analysis) is the limit we suggest. Crossing these limits is a warning to improve benchtop stability, for example by decreasing the autosampler temperature or by changing the reconstitution medium.

Choice of calibration model

The responses as obtained for the 24 calibration samples were used to establish a relationship between the concentration and response and to evaluate the goodness of fit and the lack of fit by means of analysis of variance. If a significant lack of fit is observed, measures should be taken, e.g., selection of an alternative model, application of a detector with a better performance or the application of an alternative extraction procedure.

4.2. Main validation

Precision and accuracy

The 15 measured concentrations per concentration level (triplicates from five runs) as obtained by analysing the validation samples were subjected to analysis of variance (ANOVA) to estimate the within-run precision and the between-run precision. The accuracy of the method was determined from the same results that were used for the determination of the precision. The bias of the method was calculated by comparing the mean ($\bar{x}$) of the 15 measured concentrations per concentration level with the nominal concentration (μ):

$$\text{bias}\,(\%) = (\bar{x} - \mu)/\mu \times 100$$

Recovery

For the determination of the recovery of the analytes, the mean peak heights obtained for the triplicate measurements from runs 1–5 were compared with the mean peak heights obtained from triplicate direct injections performed in the same run. The five recoveries thus obtained for each concentration level were used to calculate the mean recovery and a relative standard deviation.

Freezing and thawing stability

The three measured concentrations at each concentration level after each freezing and thawing cycle were used to calculate a mean value. These mean values were compared with the corresponding mean values obtained from the 15 precision and accuracy measurements and their 95% confidence limits (calculated from ANOVA).

Over-curve control/dilution

The 15 measured concentrations for the over-curve control samples were subjected to ANOVA. The overall mean ($\bar{x}$), the within-run precision and the between-run precision were compared with the results for the validation samples at the corresponding concentration level.

Additional to the ANOVA, a paired comparison was made between the mean measured concentration of the precision and accuracy data (normalized to a concentration of 400 ng ml^{-1}) and the mean measured concentration of the over-curve control/dilution data for each run. The null hypothesis in this comparison then is no

relative difference between results for diluted and non-diluted samples.

5. Evaluation of the results

The results of the validation experiments (as part of the pre-validation and main validation programmes presented here) are tabulated and graphically presented in Tables 2–9 and in Figs. 2–5. They will be briefly discussed below.

5.1. Pre-validation

Specificity

The specificity of the method can be illustrated by comparing the chromatograms obtained by analysing a test solution of the pure compound with the chromatograms of the independent blank plasma samples of female and male subjects. No interfering peaks were detected at the retention times of the compound of interest. The peak of the captopril–NPM adduct was well resolved and showed no interferences with endogenous or exogenous materials. No interferences from co-medications or their metabolites were observed.

Sensitivity

Under the experimental conditions described, the detection limit (defined as three times the baseline noise) was approximately 5 ng ml^{-1}. The lowest concentration of the calibration graph was 10.0 ng ml^{-1}, which was therefore the practical lower limit of quantification.

Choice of calibration model

Table 2 gives the results for the assessment of the goodness of fit/lack of fit for the total captopril in plasma assay. A graph of the data is

Table 2
Goodness of fit and lack of fit for total captopril as part of the pre-validation

Total captopril

Concentration (ng ml^{-1})	Peak height					
	X1	X2	X3	Mean	S.D.	R.S.D. (%)
10.0	561	492	511	521	35.6	6.8
25.1	1470	1586	1429	1495	81.4	5.4
50.2	2859	2956	2896	2904	49.0	1.7
100	6133	5792	6191	6039	216	3.6
251	14019	13683	13574	13759	232	1.7
502	25449	26375	29545	27123	2148	7.9
754	42088	41760	35228	39692	3869	9.7
1004	59298	57925	58589	58604	687	1.2

ANOVA table: test for goodness of fit and lack of fit

	Sum of squares	Degrees of freedom	Mean squares	F calculated	F table[a] ($\alpha = 0.05$)
Regression	$9.5 \cdot 10^9$	1	$9.5 \cdot 10^9$	2602	4.3
Residuals	$8.0 \cdot 10^7$	22	$3.6 \cdot 10^6$		
Pure error	$4.0 \cdot 10^7$	6	$6.6 \cdot 10^6$	2.62	2.74
Lack of fit	$4.0 \cdot 10^7$	16	$2.5 \cdot 10^6$		
Total	$9.5 \cdot 10^9$	23	$4.1 \cdot 10^8$		

[a] Ref. [14], p. 40.

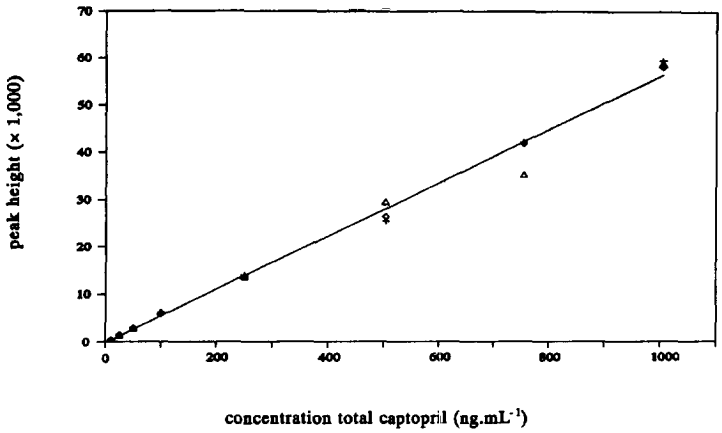

Fig. 2. Goodness of fit and lack of fit for total captopril as part of the pre-validation.

given in Fig. 2. A test for lack of fit indicated that the linear model is appropriate for establishing a relationship between the concentration and the response. No significant lack of fit was observed. The goodness of fit was highly significant. In general, for calibration graphs, correlation coefficients above 0.998 were observed during the validation experiments.

In several cases (we have seen this for other assay methods), a significant lack of fit was observed, despite the fact that it was clear that the model used was the best one. Moreover, the goodness of fit was highly significant. The lack of fit for these cases was probably caused by relatively high deviations of the actual concentration of the calibrators versus the nominal concentration (relative to the assay precision). An illustration of this is shown by paracetamol (see Fig. 3 and Table 3). A relatively high bias of nominal concentrations versus back-calculated concentrations is observed as compared with the method precision (R.S.D. $\leq$ 1.6%). Moreover, these biases (cf., mean residuals) do not follow a certain pattern, but are randomly distributed around the curve.

Despite this disadvantage in certain cases, linearity testing by means of goodness of fit/lack of fit assessment gives the analyst, in an early stage of the validation study, a very good answer to model problems, that often cannot be detected during method development stages.

A solution to problems with worse accuracy of

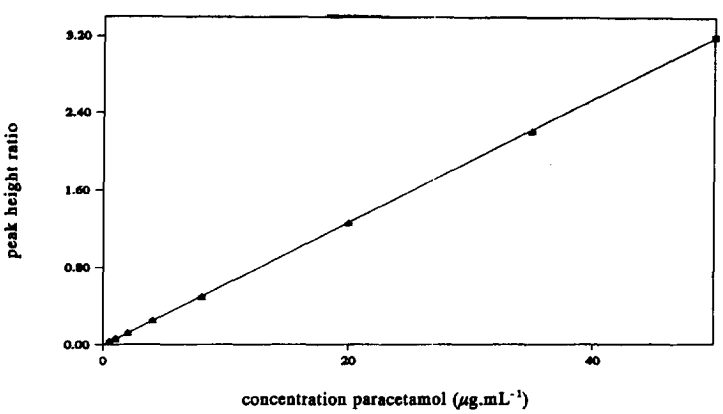

Fig. 3. Goodness of fit and lack of fit for paracetamol as part of the pre-validation.

Table 3
Goodness of fit and lack of fit for paracetamol as part of the pre-validation

Paracetamol						
Concentration (μg ml^{-1})	Peak-height ratio					
	X1	X2	X3	Mean	S.D.	R.S.D. (%)
0.50	0.0329	0.0329	0.0326	0.0328	0.0002	0.5
1.00	0.0632	0.0626	0.0612	0.0623	0.0010	1.6
2.00	0.1228	0.1234	0.1235	0.1232	0.0004	0.3
4.00	0.2480	0.2503	0.2489	0.2491	0.0012	0.5
8.00	0.5011	0.4949	0.4962	0.4974	0.0033	0.7
20.0	1.2587	1.2566	1.2650	1.2601	0.0044	0.3
35.0	2.2114	2.1983	2.2087	2.2061	0.0069	0.3
50.0	3.2130	3.1958	3.1910	3.1999	0.0116	0.4

ANOVA table: test for goodness of fit and lack of fit

	Sum of squares	Degrees of freedom	Mean squares	F calculated	F table[a] ($\alpha = 0.05$)
Regression	29.23236	1	29.23236	206292	4.3
Residuals	0.003117	22	0.000141		
Lack of fit	0.002689	6	0.000448	16.8	2.74
Pure error	0.000428	16	0.000026		
Total	29.23548	23	1.271107		

[a] Ref. [14], p. 40.

stock concentration relative to analytical imprecision during goodness of fit/lack of fit studies as presented above may be the preparation of a single stock solution and to prepare standards from this single stock solution (in this particular case only, standards are usually prepared independently). A bias in the concentration of the stock solution then has the same effect (bias) on the concentrations of the standards at all concentrations, which will not affect the slopes of the calibration model used.

Benchtop stability

The results of the experiments to investigate the stability during storage in the sample compartment of the injector are presented in Table 4. A graph for the lowest level is given in Fig. 4. Concentrations of the captopril–NPM adduct were between 90% and 110% after 30 h in the sample compartment of the injector as compared with concentrations after 0 h. No significant deterioration was observed.

Benchtop stability problems may occur owing to evaporation of reconstitution solvents. Internal standardization may eliminate these problems. If benchtop stability problems occur due to photo or thermal instability, other measures should be taken, e.g., improved storage conditions. The main validation should only be started if the benchtop stability is satisfactory (a limit of ±10% is suggested).

5.2. Main validation

Precision and accuracy

A summary of the results on precision and accuracy as derived from the measured concentrations for the validation samples is given in Table 5. A detailed table for one concentration level (25.1 ng ml^{-1}) is given in Table 6. The

Table 4
Benchtop stability (30 h) of total captopril at concentrations in the lower and higher part of the concentration range

Captopril–NPM adduct		
Time (h)	Concentration (ng ml^{-1})	
	Nominal conc.: 25.1	Nominal conc.: 754
	Peak height	Peak height
0	1858	51561
2	1886	51443
4	1723	–[a]
6	–[a]	–[a]
8	–[a]	–[a]
10	1717	48291
12	1735	48876
14	1755	48911
16	1779	48976
18	1838	50381
20	1816	50055
22	1825	52146
24	1901	51349
26	1817	49861
28	1850	53306
30	1924	51967
Slope	2.6508	51.797
Intercept	1773.2	49663
Based on regression:		
Concentration change after 30 h (%)	4.5	3.1

[a] Injection error due to pump problems.

within-run R.S.D.s were below 5.0% at all concentration levels and the between-run R.S.D.s were below 10.4% at all concentration levels.

Table 5
Summary of precision and accuracy of the analytical method for total captopril in plasma ($n = 15$)

Nominal concentration (ng ml^{-1})	Measured concentration (ng ml^{-1})	Bias (%)	Within-run R.S.D. (%)	Between-run R.S.D. (%)
10.0	10.1	1.3	4.3	10.3
25.1	26.6	5.9	4.9	7.0
251	226	−9.8	4.7	9.7
754	734	−2.6	4.3	4.0

The bias varied between −9.9% and +6.0% at all concentration levels. The summarized results indicate that Washington criteria are met.

Applying the separation of the total measurement error into its constant (bias) and random (precision) components, as suggested by Hartmann et al. [5], the between-run error at the 251 ng ml^{-1} level may be too high.

Recovery

Data on the absolute analytical recovery of paracetamol are given in Table 7. The results of the recovery experiments were satisfactory: the mean recovery of paracetamol was found to be consistent over the evaluated concentration range, and was 96.9%. The recovery of theophylline at the concentration used during the actual analysis of the plasma samples (40.0 μg ml^{-1}) was 92.6%.

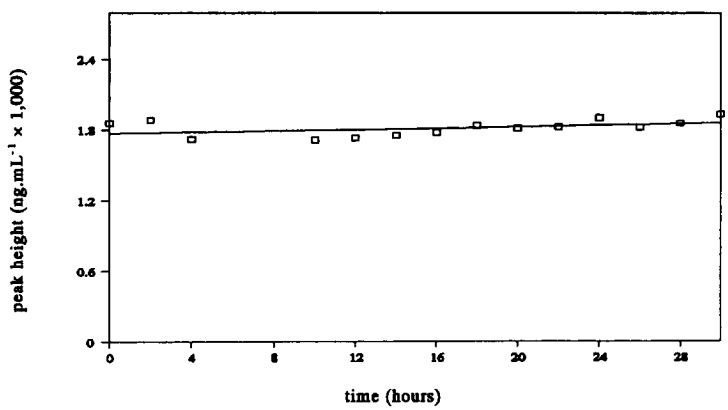

Fig. 4. Benchtop stability (30 h) of total captopril at concentrations in the lower part of the concentration range. Stability of captopril–NPM adduct in the sample compartment of the injector (ca. 25.0 ng ml^{-1}).

Table 6
Precision (within-run and between-run) and accuracy for total captopril during five analytical runs at concentrations in the lower part of the concentration range

Total captopril [nominal concentration (μ): 25.1 ng ml^{-1}]

Run No.	Date (year, month, day)	Measured concentration (ng ml^{-1})		
		X1	X2	X3
1	941027	22.8	26.7	27.0
2	941028	25.8	25.5	27.3
3	941029	25.6	25.7	26.3
4	941101	25.5	27.7	27.9
5	941102	27.8	28.1	28.7
Mean, $\bar{x}$ (ng ml^{-1})			26.6	
Bias[c], $(\bar{x} - \mu)/\mu \times 100$ (%)			5.9	
n			15	

ANOVA

	Sum of squares	Degrees of freedom	Mean squares	S.D. (ng ml^{-1})	R.S.D. (%)
Within-run[c]	17.066	10	1.707	1.3	4.9
Between-run[c]	14.018	4	3.505	1.9	7.0
95% confidence limit of $\bar{x}$ (ng ml^{-1})	$26.6 \pm 2.145^{a} \cdot 0.385^{b} = 26.6 \pm 0.825$				

[a] $t_{0.975;14}$ (Ref. [14], p. 30).
[b] Standard error calculated from within-run and between-run standard deviations.
[c] Washington criteria: at lower limit of quantification, ≤20%; at other concentrations, ≤15%.

Freezing and thawing stability

The results of the experiments to investigate the stability after repeated freezing and thawing are presented in Table 8 (25.1 ng ml^{-1} samples only). The measured concentrations after one, two, three, four and five freezing and thawing cycles did not consistently exceed the 95% confidence limits obtained from ANOVA (see precision and accuracy data in Tables 5 and 6). The results for the 25.1 ng ml^{-1} samples are also

Table 7
Recovery of paracetamol during five analytical runs for three concentrations (low, medium and high levels in the concentration range)

Paracetamol

Nominal concentration (μ) (μg ml^{-1})	Recovery (%)			No. of observations, n
	$\bar{x}$	S.D.	R.S.D.	
1.00	99.5	0.9	0.9	5
20.0	95.9	0.9	0.9	5
35.0	95.4	0.9	0.9	5
Mean	96.9	0.9	0.9	15

Table 8
Freezing and thawing stability of total captopril during five analytical runs at concentrations in the lower part of the concentration range

Total captopril (nominal concentration: 25.1 ng ml^{-1})					
Cycle No.	Date (year, month, day)	Measured concentration (ng ml^{-1})			
		X1	X2	X3	Mean
1	941027	23.4	28.9	28.6	27.0
2	941028	27.0	27.8	28.1	27.7
3	941029	26.1	27.5	27.7	27.1
4	941101	25.9	27.3	30.1	27.8
5	941102	26.4	27.4	26.9	26.9
Mean[a], $\bar{x}$ (ng ml^{-1})		27.3			
Bias, $(\bar{x}-\mu)/\mu \times 100$ (%)		8.7			
n		15			

ANOVA					
	Sum of squares	Degrees of freedom	Mean squares	S.D. (ng ml^{-1})	R.S.D. (%)
Within-run	30.4	10	3.0	1.7	6.4
Between-run	1.8	4	0.5	0.7	2.5

From precision and accuracy (see Tables 5 and 6).
[a] 95% confidence limit of $\bar{x}$ (ng ml^{-1}): $26.6 \pm (2.145 \cdot 0.385) = 26.6 \pm 0.825$.

shown in Fig. 5 (horizontal lines represent 95% confidence limits from ANOVA). No significant deterioration was observed after five freezing and thawing cycles.

Over-curve control/dilution

The results of the experiments with respect to the dilution of plasma samples are given in Table 9. The within-run and between-run R.S.D.s were

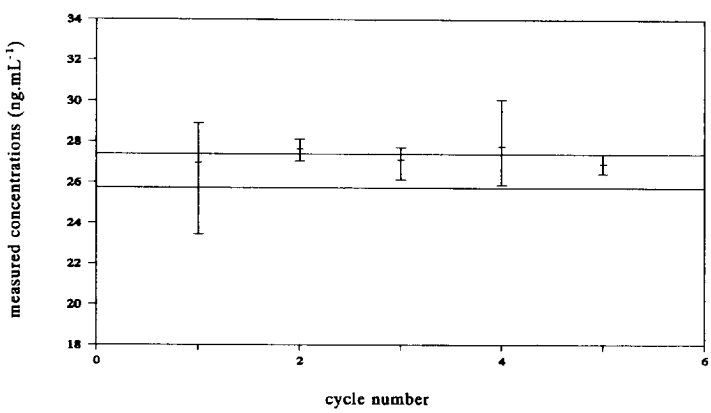

Fig. 5. Freezing and thawing stability of total captopril during five analytical runs at concentrations in the lower part of the concentration range. Nominal concentration, 25.1 ng ml^{-1}.

Table 9
Over-curve control/dilution of total captopril

Total captopril (nominal concentration 2008 ng ml^{-1}; dilution factor, 5; nominal concentration diluted, 402 ng ml^{-1})

Run No.	Date (year, month, day)	Measured concentration (ng ml^{-1})		
		X1	X2	X3
1	941027	510	536	477
2	941028	454	495	420
3	941029	367	381	379
4	941101	340	375	390
5	941102	367	382	365
Mean $\bar{x}$ (ng ml^{-1})			416	
Bias, $(\bar{x} - \mu)/\mu \times 100$ (%)			3.6	
n			15	

ANOVA

	Sum of squares	Degrees of freedom	Mean squares	S.D. (ng ml^{-1})	R.S.D. (%)
Within-run	6178	10	617.8	24.9	6.0
Between-run	47868	4	11967	109.4	26.3

Paired comparison

Analytical run	Date (year, month, day)	Precision and accuracy data	Over-curve control/ dilution data	Difference (d_i)
1	941027	367	508	−141
2	941028	392	456	−64
3	941029	365	376	−10
4	941101	347	368	−21
5	941102	340	371	−31
Mean d_i ($\bar{d}$)				−53.5
S.D.				52.8

Difference $-53.5 \pm 2.776^{a} \cdot (52.8/\sqrt{5})^{b} = -53.5 \pm 65.5$
Zero is included, the null hypothesis (no difference after dilution) can be accepted

[a] $t_{0.975;4}$ (Ref. [14], p. 30).
[b] Standard error.

6.0% and 26.3%, respectively, for a fivefold dilution. The bias was 3.6%. The within-run R.S.D. is comparable to the within-run R.S.D. obtained from the precision and accuracy experiments (Table 5). The between-run R.S.D. is significantly higher compared with the precision and accuracy results and exceeds the acceptable limit of 15%. The latter observation indicates that dilution of samples is less reliable and should be avoided as far as possible. Alternatively, a second concentration range should be adopted, if necessary.

The paired comparison between the mean measured concentration of the precision and accuracy data and the mean measured concentration of the over-curve control/dilution data

for each run is also given in Table 9. It was concluded that the null hypothesis (no significant difference between results for diluted and non-diluted samples) can be accepted.

6. Conclusions

An alternative programme for bioanalytical methods validation has been introduced. The programme, a nested, efficient and cost-effective design for chromatographic methods validation, gives the analyst the opportunity to use efficiently many experimental results for the evaluation of several performance characteristics. For recovery, freezing and thawing stability and over-curve control/dilution the same results are used as collected for the precision and accuracy validation. Only a limited number of additional experiments are performed. Nevertheless, the requirements drawn up in the Washington Conference Report are respected. Alternative validation procedures have been developed for the evaluation of the benchtop stability, the goodness/lack of fit, the accuracy and precision, the freezing and thawing stability, the recovery and over-curve control/dilution of a method from procedural and/or data processing/analysis points of view.

The recommended design combines the performance of as few independent validation experiments as possible with modern statistical methods, resulting in optimum use of information.

Acknowledgements

The authors thank Pharma Bio-Research Laboratories technicians Pauline Koopmans, Yvonne Somhorst and Monique Waalderbos for the tremendous efforts spent in developing the assay method for total captopril in human plasma and Gert Wilkens and August de Jong for the validation of the analytical method for paracetamol in plasma.

References

[1] J.R. Lang and S. Bolton, J. Pharm. Biomed. Anal., 9 (1991) 357.
[2] J.R. Lang and S. Bolton, J. Pharm. Biomed. Anal., 9 (1991) 435.
[3] A.R. Buick, M.V. Doig, S.C. Jeal, G.S. Land and R.D. McDowall, J. Pharm. Biomed. Anal., 8 (1990) 629.
[4] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman and S. Spector, J. Pharm. Sci., 81 (1992) 309.
[5] C. Hartmann, D.L. Massart and R.D. McDowall, J. Pharm. Biomed. Anal., 12 (1994) 1337.
[6] D. Dagdar, P.E. Burnett, M.G. Choc, K. Gallicano and J.W. Hooper, J. Pharm. Biomed. Anal., 13 (1995) 89.
[7] Note for Guidance, "Analytical Validation", in The Rules Governing Medicinal Products in the European Community, Vol. III, Addendum, 1990, 1.
[8] Good Laboratory Practice for Nonclinical Laboratory Studies, 21 CFR, Ch.1, Part 58, December 22, 1978, and subsequent amendments, Food and Drug Administration (FDA), Department of Health and Human Services, Rockville, MD.
[9] Good Laboratory Practice in the Testing of Chemicals, and subsequent updates, Organization for Economic Cooperation and Development (OECD), Paris, 1982.
[10] K.P. Ohman, B. Kagedal, R. Larsson and B.E. Karlberg, J. Cardiovasc. Pharmacol., 7 (1985) S20.
[11] K.L. Duchin, D.N. Mc. Kinstry, A.I. Cohen and B.H. Migdalof, Clin. Pharmacokinet., 14 (1988) 241.
[12] J. Wieling and B. Oosterhuis, Study Protocol for Assay Method, internal document, Pharma Bio-Research International, Zuidlaren, 1994.
[13] F.E. Grubbs and G. Beck, Technometrics, 14 (1972) 847.
[14] Geigy Scientific Tables—Statistics, Ciba-Geigy, Basle, 1980.