ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
14 (1996) 375–388

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

A strategy for validation of bioanalytical methods

S. Braggio*, R.J. Barnaby, P. Grossi, M. Cugola

Glaxo–Wellcome Research Laboratories, Drug Metabolism Department, Glaxo–Wellcome S.p.A., via A. Fleming, 2, 37135 Verona, Italy

Received for review 27 March 1995; revised manuscript received 9 August 1995

Keywords: Bioanalytical methods; Drug development; Validation

1. Introduction

The importance of integrating pharmacokinetics, pharmacodynamics and toxicokinetics in drug development has been emphasized in recent publications [1–4]. In this context, very often important decisions are taken which are based on data obtained from bioanalytical results. Therefore, the ability of bioanalytical methods to accurately determine the concentrations of parent drugs and their main metabolites is of prime importance. In order to design a pertinent bioanalytical assay, the analysts should initially understand the rationale of the experiment into which the analytical information will be included, what is required from the results, and what effect the bioanalytical results will have on the overall conclusion of the experiment. Once these factors are defined, the appropriate assay can be developed. Once developed, and before utilisation for real sample assay, the analyst should obtain sufficient data describing the performance of the assay in order to assure its suitability for the job in hand. This information is obtained by performing a validation study. Very briefly, a validation study involves assaying prepared samples (spiked samples) using standards prepared in the same biological matrix, and determining relevant parameters that describe the method's performance.

The most relevant parameters describing a bioanalytical method — e.g. selectivity, limit of quantification, linearity, accuracy and precision — and the criteria for their evaluation and validation have been already discussed by scientists from industry, academia and the regulatory environment [3]. However, although the overall concept is unanimously accepted within the pharmaceutical industry, there is still the necessity to standardize the procedures between different laboratories. This paper describes the procedures adopted by the author's department to validate bioanalytical methods for the research phase of drug discovery, pharmacokinetic, toxicokinetic and metabolic studies in animals and clinical studies. The design of these studies will be highly dependent on the actual development stage of the drug, and well be discussed below.

* Corresponding author.

SSDI 0731-7085(95)01644-9

2. Bioanalytical methods

The authors' department is involved in the development of a new drug from the discovery phase until postregistration clinical trials. Its role is to obtain information from the biological samples supplied by client departments. As the information needed and the number of samples can vary significantly during the whole development process, bioanalytical methods are accordingly developed, modified and revalidated. The life-cycle of a method for the quantitative assay of a new drug can be summarized as shown in Table 1.

A "preliminary" method is usually developed to support pharmacological research for the pharmacokinetic and metabolic screening of drug candidates. As only a relatively small number of samples are assayed at this stage, a preliminary validation is performed by testing only the method specificity, recovery, calibration linearity and the stability of the analytes during sample handling. After a compound has been selected for early development, the preliminary method is modified, or a new method developed for the toxicokinetic monitoring of dose-range-finding and chronic toxicity studies and absorption–distribution–metabolism–excretion (ADME) experiments conducted in animals. The validation of this "preclinical" method involves the evaluation of intra-assay precision, accuracy and linearity across an extended range of concentrations to allow the assay of high concentrations of analyte in samples from high dosage groups without the necessity of error-introducing dilution of samples. Quality control samples are stored and assayed together with unknown samples at this stage to check the daily method performance and also to give an indication of the inter-assay precision and accuracy.

A "clinical" method is developed when the project passes into phase I volunteer trials. At this stage an attempt is made to design a definitive method which would be used for the entire clinical development of the compound and therefore will already incorporate a high degree of automation. Its validation includes tests for inter-assay precision and accuracy, and long-term stability at lower concentrations. When large numbers of samples are expected from the planned phase II and III clinical trials, the method may well be further modified to include or increase automation of the method and so increase batch size.

The process for method development and the related documentation are shown in Table 2.

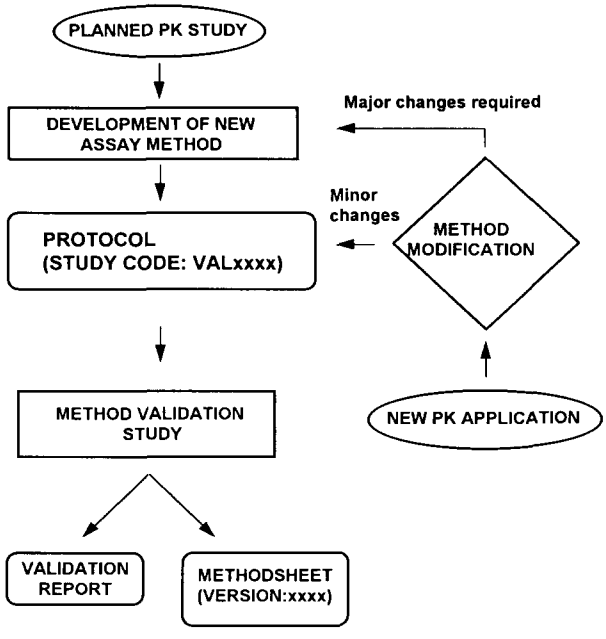
After the validation of a method, a document, given the name "method sheet", is produced and made available to all the analysts in the department. The method sheet gives exact details on method procedure and is in a recipe format. The results of each method validation are described in a specific report for submission to the regulatory authorities. It is preferable not to make any changes to a validated method because of the necessity to recheck the validation parameters. However, it is common to have to make some minor changes such as extension of the calibration range. In this case, it is our practice to perform an intra-day evaluation of precision and accuracy, and recovery and limit of quantification. If, in the case of clinical methods, the validation parame-

Table 1
Bioanalytical methods for drug development: the Glaxo–Wellcome (Verona) approach

Stage	Application	Method	Important validation parameters
Research	Preliminary pharmacokinetics and metabolism in animals	Preliminary ⇓	Specificity Recovery Stability (short-term)
Early development	Toxicokinetics ADME	Preclinical ⇓	Intra-assay (precision/accuracy) Extended calibration range QC samples
	Phase I clinical trials (volunteers)	Clinical ⇓	Inter-assay (precision/accuracy) Long-term stability
Full development	Clinical phase II, III, IV	Fully automated	Large samp. ie batch

The method should be validated for each matrix used.

Table 2
Method development and related documentation



ters from the revised method are not significantly different from those of the original method, and the validation parameters are well within the acceptable ranges, then an inter-day evaluation is not performed and the revised method is deemed validated. Otherwise, or if more drastic changes in the method are required such as in the redefinition of extration or chromatographic parameters due to an interfering metabolite, then a full validation procedure must be repeated.

3. Validation parameters and acceptance criteria

The following validation parameters should be tested and monitored, where appropriate for the type of method to be tested.

3.1. Calibration range and linearity

The calibration range will be defined by the expected concentration in the samples, and, for chromatographic assays, will be usually in the linear range of the detector for that analyte, where the determined response is directly proportional to the concentration; however it is not an absolute requisite to work in the detector linear range [4]. The calibration range should not be too wide, as the accuracy and precision will suffer at the extremes of the range. Neither should it be too narrow, as this will invariably mean, for pharmacokinetic studies, introducing dilution steps when concentrations exceed the top of the range. In our department, the usual range used between highest and lowest calibration standards is fifty to five hundred fold, utilising six to ten standards to cover this range. Another option is to include duplicate standards at four to five concentration levels if this can improve the performance of the assay. Should the range of concentrations be very large, as in toxicokinetic studies, it is recommended to validate two overlapping calibration ranges so that the samples can be included into batches containing either the high or low calibration curve. This avoids the necessity for sample dilution which can introduce unnecessary errors, especially if the dilution is high, and also avoids the use of a blank matrix.

Least-squares linear regression is normally used to mathematically define the calibration line. It is our general policy to not include the blank standard in the calibration, and not force the calibration through zero. In general, a weighting factor (usually $1/x$ or $1/x^2$) is used to avoid biasing the calibration line in favour of the high standards. This is especially important when the calibration range is wide. The actual choice of weighting is evaluated during the validation, and is used for all subsequent assays for that method. It is generally agreed that the coefficient of correlation of the calibration line should exceed 0.99 [5], however this is not a very demanding test. A better method is to determine the back-calculated (interpolated) concentrations, obtained by entering the measured response for each standard into the regression equation. The % bias for each calibration level is then calculated from the following equation

$$\%\ \text{Bias} = |\text{Nominal concentration} - \text{interpolated concentration}| \times 100/\text{Nominal concentration}$$

The total bias for each curve fitting is calculated by summing the % biases. The best fit model is the one that gives the lowest total bias and also gives the most constant % bias across the calibration range (a plot of % bias against nominal concentration is useful here). If a number of models give similar results, then the simplest one is chosen. This is the procedure followed for preclinical method validations (intra-day only). For clinical validations, the procedure for the calibration fitting model evaluation, described above, is performed on all pooled calibration data from the intra- and inter-day validations.

Calibration curves are accepted for each individual batch during validation when no one standard has a % bias greater that 20%. If the difference is larger than this value during real sample assays, then the standards can be deemed outliers, and up to two of them can be eliminated from the calibration and the linear regression repeated. The calibration is deemed unacceptable if more than two standards have to be eliminated for acceptable criteria or if less than five standards remain. During the validation procedure all standards must be used for the calibration unless there has been a clear error.

It is the general procedure in this department not to extrapolate calibration ranges to calculate concentration levels lying outside the range of the calibration, but instead samples should be reassayed after suitable dilution. In some cases, where insufficient sample is available for a second assay and the result for this sample is of crucial importance, the curve may be extrapolated up to 20% above the highest standard concentration. In addition, the fact that the result derives from an extrapolated standard curve should be mentioned in the final report.

3.2. Precision and accuracy

The precision of a method is defined [6] as the closeness of agreement between independent test results obtained under prescribed conditions. It is normally expressed in terms of the relative standard deviation (RSD) of the determined concentrations of a series of replicate assays of spiked samples. The accuracy of a method is defined [6] as the closeness of agreement between the test result and the accepted reference value. It is determined by calculating the percent difference (bias%) between the measured mean concentrations and the corresponding nominal concentrations.

The RSD and bias% define the intra-assay precision and accuracy when related to a single sample batch, and the inter-assay precision and accuracy between batches when related to multiple batches. The acceptability criteria for precision and accuracy quoted in the literature have ranged from 10 to 25%. The vast majority of reported chromatographic methods have actual values of 5–10%. It has been reported that acceptable accuracy and precision for a chromatographic method should not exceet 20% at the limit of quantification (LOQ), or 15% at other concentrations [3]. These limits are also utilized in the authors' laboratory. The acceptance criteria are relaxed slightly for radioimmunoassays which inherently are more imprecise. For these assays, the accuracy and precision should not exceed 20% and 25% at the LOQ. In extreme cases, where pharmacokinetic information (for instance elimination half-life) is required from samples containing very low concentrations, the utilized assays may not meet these requirements, and in this instance it is essential to stress the level of uncertainity in the result. Needless to say, every attempt must be made to avoid this situation especially for the advanced stages of development.

These acceptability criteria have been recently criticized by Hartmann et al. [7], who suggested that separate and much lower limits for accuracy and precision should be set, in order to ensure than an analytical result is within $\pm 15\%$ of the true result. However the limits established above are set for individual concentration levels within the calibration range and are not mean limits of the method. Typically the imprecision and bias of bioanalytical methods, especially those of very high sensitivity, will deteriorate at the lower ends of the calibration range, but will be much lower (typically less then 10%) at higher concentrations. Therefore, taken as a whole, the method will almost certainly have a much lower bias and imprecision that these limits imply. Lowering the

acceptability limits ever further would create severe practical difficulties and prolong unacceptably method development and validation.

3.3. Limit of quantification

The limit of quantification (LOQ) is defined as the lowest concentration that can be determined with acceptable accuracy and precision [8] with that particular method, and therefore is set by the lowest concentration standard used in the validation study with acceptable validation criteria. The LOQ depends on the absolute detection limit and also on the calibration range used; therefore, if the calibration range is modified, the LOQ may have to be revised. In addition, the LOQ is usually the lowest calibration standard in the actual study. However, if it is decided to exclude the lower standard(s) from the calibration, the calibration line should still not be extrapolated below the lowest standard, so the effective LOQ in this case is still the lowest standard in the calibration and not the original LOQ obtained from the validation.

Sample concentrations below the LOQ are reported as "BQL" (below the quantification limit) and not used for any pharmacokinetic calculations. They should not be set as zero.

3.4. Method linearity

The linearity of an assay method has been defined as its ability to obtain test results directly proportional to the concentration of the analyte in the samples [9]. Linearity is assessed by plotting the measured concentrations against their theoretical values using the data from the intra-day accuracy and precision evaluation. Unweighed regression line slopes should be very close to 1 with an intercept very close to 0. In general, good linearity is observed with chromatographic methods. More significant deviations from linearity can be found when the calibration curves are not linear (e.g. immunoassays).

3.5. Selectivity and specificity

Some assays are selective; others are specific [8]. Chromatographic methods are usually selective as they may detect and quantify several compounds, the analytes of interest being identified by their peak retention times. Other methods, like those based on immunoassays, must be highly specific, because it is essential that they determine only the analyte of interest and not metabolites, for example, because of the normal lack of a chromatographic separation technique. However, even immunoassays cannot be said to be completely specific, in the true scientific sense, because of some possible cross-reactivity or non-specific binding (see Section 5).

Evidence that the assay does not suffer interference from endogenous compounds is obtained by assaying control blank samples from at least six individuals [3] for clinical assays and at least three individuals for preclinical assays, preferably both before and after food intake. Three individuals are usually sufficient for animal samples because animals bred and housed under strictly controlled conditions are very homogenous (same strain, age, weight and diet), unlike humans. However, if a clear, measurable interference is present, six individual animal samples should be tested to estimate the percentage interference compared to the lowest calibration standard. If the clinical method is likely to be established in different countries in the world then it becomes important to test more control samples of the population(s) in question. However, experience has shown that there is no guarantee that any study sample will not contain an interfering compound not observed during method validation. Therefore it is good practice to take a predose sample from each individual generating study samples. Frequently some slight inteference is present, especially working with extremely low analyte concentrations. We set the criterion that the intercept of the calibration regression line (using the guidelines for method regression already stated) should not exceed 20% of the signal at the LOQ. This criterion therefore will also dictate the LOQ itself.

Preferably, selectivity against metabolites should also be demonstrated. In the past this was

generally only possible in advanced phases of drug development; however with the recent availability of in vitro metabolic techniques, interference from probable in vivo metabolites can be checked for during both preclinical and clinical method validation procedures. If this is not possible, then this uncertainity of the selectivity of the assay should be mentioned in the final validation report. In later stages of development it is important to show selectivity against co-administered drugs: blank matrix should be fortified with the drugs in question and their metabolites, if possible at the maximum expected concentration for clinical samples containing these drugs.

The specificity of a bioanalytical method should preferably be finally established for real samples by using a spectroscopic method such as mass spectrometry to prove that the peak of interest corresponds to the analyte of interest and that the peak is pure. Diode-array detectors may provide information on the purity of the peak. An alternative to mass spectrometry is to perform the sample analysis by a second chromatographic method with different selectivity and which has previously been validated.

3.6. Enantioselectivity

The topic of enantioselectivity in bioanalytical methods has received much attention [10] and will not be discussed in detail here. Suffice to say that when a single-enantiomer drug is developed, a chiral method has to be used to show for each species and for each route of administration that there is no racemization in vivo. This method does not have to be fully validated as the end result is solely qualitative. If in vivo racemization does occur or if the racemic drug is being developed, then a chiral method must be developed and validated. Validation of enantiomeric methods is more problematical as two compounds, quite often eluting very close together, have to be quantified. In this case it would be preferable to validate, for each concentration level, a series of enantiomeric ratios, but this would increase dramatically the number of assays performed in a validation. A compromise has to be made and this is discussed further in the literature [11–15].

3.7. Recovery

The recovery is a measure of the efficiency of the method in detecting all the analyte of interest present in the original sample. It is measured in this laboratory by comparing the response of extracted samples spiked before extraction with the response of extracted blank matrix samples spiked just before injection or response determination. This is performed rather than a comparison with pure unextracted standards in order to compensate any effect the matrix may have in the signal response. If internal standards are used, the recovery of the internal standard should also be checked using the test analyte as internal standard. In this case the actual internal standard used is spiked after extraction. Recovery should be as high as possible, but it is more important that it is constant within the calibration range. Therefore, in order to check this, the recovery is evaluated close to the extremes of the calibration range and at an intermediate concentration. A general minimum target in this department is 70%, although this may be more difficult to achieve with some techniques such as Asted.

3.8. Stability after sample processing

It is extremely important to perform a stability study of the analyte in biological fluids as soon as possible in the lifetime of the project in order to obtain information concerning the conditions and times of sample storage so that sample integrity before assay is assured. The design of these studies is described later. This section deals with an equally important factor; that of stability after sample extraction. (It is assumed that the stability during assay is high, as this will naturally be reflected in the recoveries obtained.) This information should be generated very early in a validation study as it may dictate the conditions under which the assay is performed (e.g. use of cooled autosamplers) and can determine the batch size and perhaps influence the method itself. This information, generated by reinjecting an extracted standard and a real sample over a period of time, is extremely useful in determining how long the extracted sample can be stored after, for instance, an overnight instrument failure.

An important consideration to take, once the stability after assay has been established, is to repeat the tests with real samples, as sometimes unstable metabolites, especially phase II conjugates, may convert back to the parent compound on storage or after the sample assay. The stability during assay may also be important in this situation, especially if extremes of pH are used.

4. The validation study

Before describing the actual validation design for each development stage of a drug, a number of important practical points should be considered.

4.1. Biological matrix

Each biological matrix from each species should be validated. Also, if at all possible, during the actual sample assay, the same matrix as the samples should be used for preparing standards and quality controls. The latter is sometimes not always possible because of limited availability, for example in mouse plasma assays, and in this situation the validation procedure can incorporate a calibration in a substitute matrix to check its suitability for calibration standard preparation as long as the method has already been validated in this matrix [3]. Also in the authors' department, horse plasma is sometimes used as a substitute for human plasma (where high volumes are required) to prepare calibration standards, but not quality control samples, unless long-term stability data have also been obtained in horse plasma. In order to do this, an intra-day validation in horse plasma is performed alongside that in human plasma.

Finally in the method report, it is important to document the anticoagulant used, the anaesthetic used to obtain the control matrix, and the exact materials for vials, etc. These are all minor details, but which sometimes become very important when a small change is made.

4.2. Analytical working standard

Before contemplating a validation study, it is essential to have a fully characterized analytical working standard with known purity and exact conversion factor to calculate the concentration of free acid or base if the analyte is a salt. Also it is essential to have sufficient information on the stability and storage conditions of the analyte as a solid and in solution in order to complete a validation study without problems of changing analyte composition [4]. Before starting a validation study, any precautions regarding standard weighing in particular should be known.

4.3. Preparation of calibration standards, spiked samples and quality controls

The validation procedure is performed by assaying a blank matrix fortified with the analyte (spiked samples), using calibration standards prepared in the same matrix. The volume of the solution added to the matrix in general should be small when compared to the volume of the matrix, but the essential point is that the calibration samples, spiked samples and, eventually the real samples are treated all in the same manner. One important point is that the volume added to the blank matrix to prepare quality controls should be less than 5% of the matrix volume, and the spiking solution should contain the minimum quantity of solvent to simulate as much as possible a real sample.

Spiked samples and calibration standards are always prepared from separate weighings. Before starting the intra-day validation, two stock solutions (A and B) are prepared, and the concentration of analyte present determined using the conditions of the method without extraction. The two solutions used should contain concentrations that are within 5% of each other. If this condition is not met, further solutions are prepared, outliers identified and discarded. Solution A is used to prepare calibration standards for the intra-day validation and solution B is used to prepare spiked samples for both intra-day and inter-day validations. It is important that this control is performed because larger differences in standard

weighings will falsify the bias determination in the intra-day validation. In order to introduce some realistic variability into the inter-day validation runs, to simulate the assay of real samples, calibration standards are prepared freshly each day [16].

Likewise, quality control samples are prepared from a different weighing (solution C) and the actual concentration determined on the day of preparation.

4.4. Batch size and composition

Before embarking upon a validation study, the intended batch size must be known so that the batch size can be tested in the validation. Thereafter the maximum batch size possible with that method is that tested in the validation procedure. If the intended batch size is very large, it may be necessary to prepare additional samples to those suggested below. The samples (calibration samples, quality controls and spiked samples) should be injected in a random fashion to avoid bias if the assay drifts during an assay batch.

4.5. Analytical instrumentation

We perform preclinical and clinical method validations in the spirit of good laboratory practice (GLP) regulations. That is, the study is performed in GLP but there is no quality assurance audit. Therefore, before commencing a validation, the analytical instrumenation to be used is checked for correct performance, If the method is chromatographic, checks are made of the accuracy of the pump flow, repeatability and linearity of the autoinjector, and the correct function of the detector in terms of signal and noise. Finally all pipettes and balances to be used in the validation are calibrated. The details of these checks are recorded and archived with the validation data and final report.

4.6. Validation of preliminary methods

A preliminary assay method is one which will be used for pharmacokinetic studies on primary research (discovery) compounds. Drug metabolism and pharmacokinetic studies have become an essential factor in research projects as it is preferable to optimize pharmacokinetics of compounds as early as possible before they reach a development phase where pharmacokinetic problems become more difficult, if not impossible, to resolve. Indeed, quite often the final selection of compound(s) to progress into exploratory development is based on pharmacokinetic properties. At this stage of a project an essential element under consideration is rapidity of turnover of results and hence the minimum of validation is performed. However, the analyst has to be very careful and consider all the possible errors in the results reported, especially in a situation where pharmacokinetic data dictate the direction of the project. The main parameters to determine for the validity of a preliminary method are the recovery, specificity, detection limit, linearity of the calibration, stability in biological matrices and stability after extraction. However, if the bioanalytical data are to be used to decide between several compounds for progression to exploratory development than the analyst should consider the need to obtain some precision and accuracy data, before proceeding with the studies.

The criteria for recovery and selectivity are as described in the relevant sections above. A test of recovery can be performed by comparison of a spiked, extracted standard and a pure standard, and should be performed at the two extremes of the expected calibration range. The specificity should be checked on at least two different batches of preferably fresh matrix. The detection limit at this stage is simply the lowest calibration sample used, and cannot be given the term "limit of quantification" because of the absence of precision and accuracy data. As above, the calibration line should have a regression coefficient of correlation of more than 0.990 and should preferably be more than 0.995. The curve should be prepared at least twice to check for reproducibility.

It is important that a brief stability check is made in the biological fluid of interest (blood, urine, etc.) at 37°C for at least 4 h to ensure that an in vivo experiment is possible and that there are no gross problems of analyte instability in the sample. Also, a check on stability after sample

clean-up is made to determine if the sample batch size is limited or if there are problems reinjecting the day after sample preparation, after storage at ambient temperature. As a standard procedure for the assay of new compounds, fresh blood and plasma samples are always stored temporarily in ice and then stored, pending assay, at −80°C. Urine and feces samples are collected into containers kept at below −20°C and likewise finally stored at −80°C. As a further control of the stability of the samples and the validity of real sample assay, quality controls are prepared on the day of sampling, stored with the samples and assayed at the same time as the samples. At least two quality controls are assayed with every batch of samples. They should be prepared at a single concentration level, near the low-to-middle part of the calibration range, and both measured concentrations should be within 25% of the theoretical concentration.

4.7. Validation of preclinical methods

Before the clinical trials, the exploratory development involves an intense program of experiments to be carried out with the new drug candidate in animals. The assay of drug concentrations for pharmacokinetic (ADME studies) and toxicokinetic (safety studies) evaluation is an essential part of these preclinical studies which are performed under GLP regulations. At this stage, a preliminary method has already been developed with a good idea of the LOQ, and quite often the method is used or modified slightly for use as a preclinical method.

A preclinical method validation involves an evaluation of intra-day accuracy and precision, and the following samples are prepared and assayed in a single batch.

(1) An appropriate set of calibration samples in order to build the calibration curve.

(2) Sixfold replicate spiked samples at not less than three concentration levels (the expected LOQ, an intermediate and the highest concentration of calibration samples, respectively). More levels can be added to make up the sample batch size to that anticipated.

(3) A set of quality controls

A recovery check is performed in duplicate at three concentration levels, one close to the lower end, one in the middle and one towards the upper end of the calibration range.

Blank samples from at least three different animals (see Section 3.5) are analyzed to demonstrate the selectivity of the method against endogenous compounds.

Quality control samples are prepared, on the day of the intra-day validation, at a medium-to-low concentration in the calibration range and are used at this stage as a further check of the correct preparation of the standard solutions for the validation study. At least four quality control samples should be assayed. As a guide line, the nominal concentration should be within the 95% confidence interval around the determined mean from intra-day spiked samples. Subsequently, for real sample assay, at least three quality controls should be assayed in each batch and at least two of them should be within 20% of the theoretical value.

4.8. Validation of clinical methods

At this stage, quite often the method undergoes substantial modification, the major reasons being a required increase in sensitivity, precision and accuracy and a strong consideration of sample turnover and hence automation. It is possible that during the lifetime of the clinical trials, new targets of sample turnover may have to met, or new technology becomes available to further improve the method. In this case the original clinical method may be further automated (see Table 1); however there is no distinction between clinical and automated methods as far as the validation procedure is concerned.

Before use in clinical studies, a comprehensive validation procedure is performed to monitor both intra- and inter-day accuracy and precision, and also the results of quality control samples assayed over a number of days.

The following samples are assayed in the same batch for intra-assay validation.

(1) An appropriate set of calibration samples to build the calibration curve.

(2) At least sixfold replicate spiked samples at the same concentration as each of the calibration samples.

(3) A set of 12 quality controls at three concentration levels.

The following samples are prepared on at least four sample batches and assayed on different days for the inter-assay validation.

(1) An appropriate set of calibration samples to build the calibration curve.

(2) at least **triplicate** spiked samples at the same concentrations as the calibration standards.

(3) A set of six quality controls at three concentration levels.

(4) Drug-free matrix samples taken from at least six different subjects.

In order to calculate the inter-day precision, the coefficient of variation of the means is determined from the four or more mean concentrations determined on individual days at each level so as not to confuse inter-day and intra-day variability. Alternatively, the analysis of variance can be used to separate the inter-day variation from the intra-day variation [17].

Quality control samples are prepared on the day of the intra-day validation by obtaining a pool in the matrix of interest at each of three concentrations and then dividing the pools into aliquots of the same volume used in the method. The concentrations are selected (they should be the same as those of the calibration standards) so as to be near the lower, middle and upper part of the calibration range, respectively. At least four replicates at each concentration level are assayed immediately to determine the actual mean concentration. As in the preclinical validation, the nominal concentrations should be within the 95% confidence interval around the determined mean values. The quality controls are subsequently stored under the same conditions as the real study samples will be, and assayed with every batch of the inter-day validation study samples.

The assay of quality controls during the inter-day validation serves as a check that the acceptance criteria set for the quality controls at each concentration level are valid. These criteria are based on the RSD values obtained from spiked sample replicates during the intra-day validation and will subsequently be used for real sample assay batches. They are calculated as shown in Table 3. These limits shown are set so as to guarantee the unacceptance of data when a method is "out of control" but do not invalidate date produced from a very accurate and precise method when the data are acceptable for the purposes of the study.

Therefore during the inter-day validation and the actual studies, at least four of the six quality controls included in a sample batch and at least one at each concentration level should be within the acceptable interval around the actual mean concentration. These criteria seem to have been accepted by most laboratories [3,8,18]. If, during the validation these criteria are not met, the acceptance criteria can be altered if the discrepancy is small and the acceptable interval does not exceed 20% or 25% for the radioimmunoassays; however it would be more advisable to seek ways of improving the method performance.

4.9. *Method transfer*

If a method is transferred from analyst to analyst, or a different type of instrument is used, some checks should be made to ensure that the method performance and accuracy are not altered. The method should be checked by performing, on at least two different days, a quality control assay using the normal number of calibration samples. Acceptable quality control results, using the criteria described previously, should be obtained on both occasions and without obvious bias compared to the initial determined mean concentration.

Table 3
Calculation of acceptance criteria set for the quality controls

RSD of quality control concentration, intra-day validation (%)	Acceptable interval around mean value, chromatographic method (%)	Acceptable interval around mean value, RIA (%)
< 5	±10	±15
5–10	±2 RSD	±20
> 10	±20	±25

If a method is to be used by a different laboratory, such as a contract house, the validation procedure for that type of method must be performed together with a series of blind quality controls prepared by the sponsor, or, alternatively, real samples already assayed by the sponsor. Again the previously described acceptance criteria apply.

4.10. Within-study monitoring

The validation of a method should not finish with the formal prestudy validation study, but should continue during the actual studies in which it is used. Calibration parameters, selectivity, and above all, quality controls should be monitored during a large number of assay batches to check for trends and drifts of assay performance. Quality controls are prepared and assayed as in the validation procedure and the same acceptance criteria apply. Quality control charts such as the Stewhart or cusum plots [3,4,17] are useful for long-term method performance evaluation.

5. Validation of radioimmunoassays

Radioimmunoassays (RIAs) are usually developed during the clinical phase of a project in order to improve sensitivity and sample throughput for assay in human plasma. The validation design of a radioimmunoassay should be exactly the same as for a chromatographic method, with slightly more lenient acceptability criteria (see Section 3.2 and Section 4.8); nevertheless there are special requirements regarding specificity to be considered when validating an RIA method. The following additional tests are performed.

5.1. Cross-reactions

When determining the specificity of a new RIA technique, it is necessary to include information on the cross-reactivity of the particular antiserum employed with substances structurally related to the primary antigen (the analyte). Limited cross-reactions (1% or lower) of at least four to eight different compounds, known metabolites if available, are often used as an argument to claim "specificity" for the RIA. However, it is important to realize that such a limited number of compounds tested is usually not enough to be sure of a specific interaction, and the other tests should be performed, as described below.

5.2. Parallelism

Potential cross-reacting metabolites of a drug in a biological sample would be expected to have a different affinity for the anti-serum compared to the parent drug. This would lead to a different dose–response curve relative to that obtained with the parent drug.

A plasma (or urine) sample from at least one subject having received the compound under investigation should be serially diluted with blank plasma (or urine) for at least five consecutive dilutions, ensuring that the concentration of the sample after the last dilution falls in the calibration range. These samples are then assayed under normal conditions. If cross-reacting metabolites, in addition to the parent drug, are present in the sample, the concentrations of the samples corrected for the dilutions would be different.

5.3. Cross-validation

A direct comparison of RIA measurements with those performed by an independent assay method in the relevant biological fluid should be sought whenever such a method exists and is accessible to the investigator. This comparison should preferably be performed blind, and involve a reasonable number of samples (20–40) spanning the whole range of concentrations of the calibrations range and sample time points, at all possible. Classical statistical methods can be used to detect significant differences [19,20].

6. Stability in biological fluids

Knowledge of the stability of an analyte is essential to be able to guarantee the integrity of the sample so that both validation procedures and real sample assays can be performed and give

meaningful results. Instability can occur during any step from the sampling stage to the final assay. Initial stability studies determining the short-term stability have already been dealt with in the validation of preliminary methods (Section 4.6). Further studies may be necessary, before commencing the in vivo experiments, to define exactly the collection and storage conditions for each matrix. The analyte stability should be studied in blood or plasma at room temperature for up to 6 h and urine or feces for at least 24 h [4]. This section deals with the design and operation of longer-term stability studies.

6.1. Preclinical stability studies

As soon as a validated preclinical method is available, stability studies at −20 and −80°C should be initiated. At this stage, stability in animal biological fluids is studied but it may be convenient to also include a study in human fluids even though the method will probably not have been validated for these matrices, to give some early indication of any differences between human and animal fluids. Stability tests should be conducted in all biological fluids from all species that will be studied in the future. The concentration levels studied should reflect the likely concentration to be found during ADME and toxicokinetic studies. At least two concentration levels should be studied: a lower concentration near the lower end of the calibration and one towards the higher end. If a dilution step is anticipated for high concentrations, then the quality control samples should be prepared at this high concentration and the same dilution should be performed.

Samples should be prepared as for the quality control samples, i.e. a pool for each concentration divided up into the aliquot volumes used in the method. Again, as with the quality control assay, at least four samples at each concentration are assayed immediately in order to obtain the actual mean starting concentration, and then at every sampling time at least three aliquots per concentration are assayed. Enough samples should be prepared to determine the stability at approximately ten sampling times up to 1 year of storage. Of course, if instability is observed after the early sampling times at −20°C, then only the study at −80°C is continued. At this stage, stability data for a few months' storage are usually sufficient, but longer-term data may be of use and it does not require a great deal more resources to continue the study up to 1 year.

6.2. Clinical stability studies

It is more probably at this stage of the development of a drug that biological samples may need to be stored for long periods, and stability data up to 1 year and possibly for longer are essential. Another important consideration at this stage is the expected concentration range in human plasma samples, which is usually considerably less than that found in preclinical studies. Hence these studies usually cannot start before the clinical method has been developed and validated. Again, at least two concentrations should be studied and the preparation and assay performed as above.

6.3. Freeze–thaw cycles

Instability can occur during the process of thawing and refreezing, and a study of the stability after repeated freeze–thaw cycles for each matrix of interest, is important, when there is a necessity to reassay samples. These studies can normally be incorporated into normal long-term stability studies. The samples are prepared as for other stability studies and at the same concentrations. Aliquots are assayed on the day of preparation and the rest are frozen. After at least 24 h freezing, all samples are brought to room temperature without heating; some are assayed and the rest are refrozen again. This is repeated for a further two cycles. It is essential to test some real samples under the same conditions; for the reasons described below.

6.4. Real sample stability

The aim of the above studies is to obtain information on the stability of the analyte in real study samples in the matrix of interest, and to establish storage conditions and lengths of storage. However, of course the stability studies give us infor-

mation of the stability not in real samples but in the blank matrix. These data can usually be extrapolated to the real sample case, but care must be exercised as this assumption can lead to problems. In particular, labile metabolites such as *N*-oxides or phase II conjugates such as glucuronides or glutathiones can be sufficiently unstable to reconvert back to the parent compound on storage or thawing, and this can lead to falsely high determined concentrations. This should be suspected as soon as evidence of the structure of the metabolites is obtained, but in our opinion it is best to check for this problem as soon as possible in the first studies with each species matrix. Real samples and quality controls or calibration standards should be reassayed after leaving them for 1 day and 1 week at ambient temperature. If real sample stability appears to be a problem when compared to quality control or calibration standard stability, then a complete stability study as described above should be performed on pooled real samples.

6.5. Data analysis and instability evaluation

In the authors' department, a procedure written with RS/1 (BBN Software Products Corporation) software is used for stability data analysis. A first-order kinetic degradation process of the analyte is assumed. The logarithm of the generated response signal, i.e. peak area or height values, is plotted against storage times, and linear regression of the date is performed. The results are plotted together with a calculated 95% lower confidence limit curve. The drug is considered to be stable in the biological fluid as long as (a) the concentration estimates from the regression line are higher than 95% of the initial concentration; or (b) the 95% lower confidence limit curve is higher than 90% of the initial concentration.

The initial concentration is determined by the intercept of the regression line with the response axis. Therefore, the time points at which the above two conditions are no longer true are calculated, and the lowest of the two times is taken as being the stability of the analyte.

7. Conclusions

Bioanalysis and the production of pharmacokinetic, toxicokinetic and metabolic data play a fundamental role in pharmaceutical research and development; therefore the data must be produced to acceptable scientific standards. For this reason and the need to satisfy regulatory authority requirements, all bioanalytical methods should be properly validated. In this paper, validation procedures used in the authors' department have been described; most of these procedures are consistent with those in the literature. It is hoped that these validation guidelines have taken into account the statistical arguments described in the literature but also have regard to the practicalities of performing bioanalytical method validations for the pharmaceutical industry in this highly competitive era and that they aid further standardization in this field.

Acknowledgements

The authors' wish to thank Dr. A Pateman of Drug Metabolism, Glaxo, UK, and Dr. K Selinger of Clinical Pharmacology, Glaxo, USA for their guidance in the preparation of this paper.

References

[1] C.C. Peck, J. Clin. Pharmacol., 34 (1994) 11–119.
[2] D.A. Smith, K. Beaumont, N.J. Cussans, M.J. Humphrey, S.G. Jezequel, D.J. Rance, D.A. Stopher and D.K. Walker, Xenobiotica, 22 (1992) 195–205.
[3] 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, Eur. J. Drug Metab. Pharmacokin., 16 (1991) 249–255.
[4] A.R. Buick, M.V. Doig, S.C. Jeal, G.S. Land and R.D. McDowall, J. Pharm. Biomed. Anal., 8 (1990) 629–637.
[5] A.G. Causey, H.M. Hills and L.J. Phillips, J. Pharm. Biomed. Anal., 8 (1990) 625–628.
[6] International Organisation for standardisation in Statistical Methods, ISO Standards Handbook 3, 3rd edn.,1989.
[7] C. Hartmann, D.L. Massart and R.D. McDowall, J. Pharm. Biomed. Anal., 12 (1994) 1337–1343.

[8] H.T. Karnes, G. Shiu and V.P. Shah, Pharm. Res., 8 (1991) 421–426.
[9] C.P.M.P Notes of Guidance: Analytical validation, The Rules governing Medicinal Products in the European Community, Addendum 3, European Communities Commission Luxembourg.
[10] F. Jamali, Pharmacology, 32 (1992) 930–934.
[11] R.W. Abbot, G.D. Allen and G. Rhodes, in E. Reids, H.M. Hill and I.D. Wilson (Eds.), Biofluids and Tissue Analysis for Drugs, Vol 23 The Royal Society of Chemistry, 1994, pp 255–256.
[12] R.R. Walters and C.Y. Hsu, Chirality, 6 (2) (1994) 105–115.
[13] A. Marzo, Boll. Chimi. Farma., 132 (8) (1993) 267–271.
[14] H. Fieger, J. Iredale and I.W. Wainer, Chirality, 5 (2) (1993) 65–70.
[15] T.J. Wozniak, R.J. Bopp and E.C. Jensen, J. Pharm. Biomed. Anal., 9 (1991) 363–382.
[16] J.R. Lang and S. Bolton, J. Pharm. Biomed. Anal., 9 (1991) 357–361.
[17] J.R. Lang and S. Bolton, J. Pharm. Biomed. Anal., 9 (1991) 435–442.
[18] A.C. Cartwright, Drug Inf. J., 25 (1991) 471–482.
[19] B.S. Chakraborty, E.M. Hawes, Rosenthaler and K.K. Midha, Ther. Drug Monit., 9 (1987) 227–235.
[20] R. Rahmani, P. Gil, M. Martin, A. Durand, J. Barbet and J.P. Cano, J. Pharm. Biomed. Anal., 1 (1983) 301–309.