COMMUNITY REFERENCE LABORATORIES RESIDUES (CRLs)

20/1/2010

GUIDELINES FOR THE VALIDATION OF SCREENING METHODS FOR RESIDUES OF VETERINARY MEDICINES

(INITIAL VALIDATION AND TRANSFER)

Table of Contents

1. Scope

This guideline document supplements Commission Decision 2002/657/EC [1] regarding the validation of screening methods. The guideline covers two distinct phases in the validation process: the initial validation of screening methods in the originating laboratory and the shortened or 'abridged' validation of these methods in the receiving laboratory following their transfer to that laboratory. The objectives of this guideline document are to define:

- the minimum requirements to be fulfilled by the initial validation (in the 'originator' laboratory);
- criteria which are necessary to determine whether screening methods can be transferred to another laboratory and under which conditions;
- the minimum requirements to be fulfilled by the abridged validation (in the 'receptor' laboratory);

This guideline document includes:

- an 'initial validation' protocol for demonstrating performance characteristics for newly developed or introduced screening methods;
- a description of the conditions under which methods developed and validated according to Commission Decision 2002/657/EC [1] in one laboratory (hereafter referred to as the 'originator' laboratory) can be transferred to a 'receptor' laboratory and the abridged validation necessary to demonstrate that the receptor laboratory is able to apply the transferred method correctly; and
- recommendations on routine quality control (continuous verification) for screening methods.

2. Definitions

2.1. Regulatory/Action Limit for validation purposes

For the purpose of validation of analytical methods in the laboratories it is understood that the Regulatory Limit for authorised veterinary medicinal products in the EU is the Maximum Residue Limit (MRL) as defined in Regulation (EC) No 470/2009 [2] (repealing Council Regulation (EEC) No 2377/90 [3]). Also, the Regulatory Limit for certain prohibited or unauthorised analytes is the Minimum Required Performance Limit (MRPL) *or* the Reference Point for Action (RPA) as defined in Article 4 of Commission Decision 2002/657/EC [1] and Article 2 of Commission Decision 2005/34/EC [4] and Articles 18/19 of Council Regulation (EEC) No 470/2009 [2].

2.2. Screening Target Concentration

The Screening Target Concentration is the concentration at which a screening test categorises the sample as “Screen Positive” (potentially non-compliant) and triggers a confirmatory test.

1- For authorised analytes, the Screening Target Concentration is at or below the Regulatory Limit (MRL)

(and should preferably be set at one half of the MRL wherever possible).

2- For prohibited & unauthorised analytes, the Screening Target Concentration must be at or less than the MRPL (or RPA).

3- For analytes for which Maximum Residue Limits (MRLs) have not been established according to Council Regulation (EC) No 470/2009 [2], Screening Target Concentration should wherever possible be at or less than the recommended concentrations as described in the CRL Guidance Paper of 7 December 2007 [5].

The further the Screening Target Concentration is below the Regulatory Limit, the lower the probability of obtaining a false compliant (i.e. false-negative) result in samples containing the drug at the Regulatory Limit.

2.3. Detection capability CCβ

Detection capability (CCβ) is defined in point 1.12. of the Annex to Commission Decision 2002/657/EC [1]. CCβ is the smallest content of the analyte that may be detected, identified and/or quantified in a sample with an error probability of β. The β error is the probability that the tested sample is truly non-compliant even though a compliant measurement has been obtained. **For screening tests the β error (i.e. false compliant rate) should be < 5%.**

In the case of analytes for which no Regulatory Limit has been established, CCβ is the lowest concentration at which a method is able to detect truly contaminated samples with a statistical certainty of 1 – β. **In this case, CCβ must be as low as possible or lower than Recommended Concentrations when exist [5].**

In the case of analytes with an established Regulatory Limit, CCβ is the concentration at which the method is able to detect permitted limit concentrations with a statistical certainty of 1 – β. **CCβ is the concentration at which only ≤ 5 % false compliant results remain. In this case, CCβ must be less than or equal to the Regulatory Limit.**

2.4. Cut-Off Level

The Cut-Off Level is the response or signal from a screening test which indicates that a sample contains an analyte at or above the Screening Target Concentration. If the Cut-Off Level is exceeded a subsequent confirmatory test is carried out. During the initial validation process the Cut-Off Level may be established through analysis of matrix blank samples and replicates of those same samples spiked (fortified) at the Screening Target Concentration. Two examples of how Cut-Off levels could be established are given in Annexes I and II.

2.5. "Negative Control" (Blank matrix) Samples.

These are samples from animals of known history which have not been exposed to the substance in question. If samples from such animals are not available, samples which have been previously confirmed as compliant and not containing residues of the substance of interest by suitably sensitive physicochemical tests would also be acceptable.

2.6. “Screen Positive Control" Samples.

These are Negative Control Samples that are fortified (spiked) with the test analyte at the Screening Target Concentration. They may, however also be incurred-positive samples (i.e. samples taken from animals which have been treated with the substance in question) or Certified Reference Materials. When Screen Positive Control samples are run in the screening test, they should be classified as 'screening positive' if the test is performing correctly.

2.7. Test Matrix

The test matrix is the tissue or fluid submitted for analysis e.g. liver, kidney, urine, muscle, honey, milk. The test matrix is specified in the Standard Operating Procedure (SOP) of the screening method. For example the test matrix may be "bovine muscle" or may be just "muscle". In the former case the screening method has been shown to be applicable to bovine muscle. In the latter case, the method has been shown to be applicable to muscle from several species with no significant difference in the test responses between the different species. The applicable species should be specified in the SOP.

2.8. Standard Operating Procedure (SOP) for the method

The method SOP should be drafted *before* any initial or abridged validation is carried out and should cover those points listed in section 5.4.4. of ISO 17025:2005. During the development of an analytical method, the scope of the method has to be determined (i.e. the analytes which can be detected, the matrices which can be tested etc). During method development, all relevant parameters of the method have to be optimised and critical points (e.g. temperature, pH) of the method identified and controlled.

When a method is transferred to a receptor laboratory, this laboratory should draw up its own SOP which should correspond with the SOP of the originator laboratory. Concerning physico-chemical methods, the equipments could be supplied by different manufacturers in the originator laboratory and in the receptor laboratory; therefore the performance could be significantly different. In the receptor

laboratory, the operator will be free to adapt the conditions with the only objective to achieve the same performance criteria.

2.9. Initial validation

The validation procedure applied to a newly developed analytical method in the originator laboratory which demonstrates that the method is fit for purpose.

2.10. Abridged validation

The shortened validation procedure applied in the receptor laboratory to a method which has been previously validated in the originator laboratory. The abridged validation procedure should allow the receptor laboratory to demonstrate that the method will work reliably in that laboratory and is fit for purpose.

2.11. Screening Methods

Screening methods are defined in Commission Decision 2002/657 [1] as "methods used to detect the presence of an analyte or class of analytes at the level of interest. These methods have the capability for a high sample throughput and are used to sift large numbers of samples for potential non-compliant results. They are specially designed to avoid false compliant results".

The “Level of interest” is usually either the Regulatory Limit (MRL, MRPL) or an “Action Level/Limit”. (see section 2.2).

3. Screening Method classification

Screening methods can be classified either according to the principle of detection or according to whether they are qualitative or (semi-)quantitative.

3.1. Classification by detection principle

Biological methods detect cellular responses to analytes (e.g. oestrogenic effect, inhibition of bacterial growth, cellular effect, hormonal effect). These methods are not selective and can cover several chemical classes of active analytes (e.g. hormones, antimicrobials). They do not allow the identification of individual analytes.

Biochemical methods detect molecular interactions (e.g. antigens, proteins) between analytes and antibodies or receptor proteins (ELISA, RIA, …). Chemical labelling of either the analyte or antibody/receptor allows the interaction to be monitored and measured. These methods are either selective for a family of analytes having related molecular structures or are sometimes analyte-specific.

Physicochemical methods distinguish the chemical structure and molecular characteristics of analytes by separation of molecules (e.g. TLC, GC, HPLC) and the detection of signals related to molecular characteristics (e.g. UV- DAD, Visible, Fluo, FID, ECD, MS, tandem MS, trap MS, ToF MS, other hybrid MS). They are able to distinguish between similar molecular structures and allow the simultaneous analysis of several analytes.

3.2. Classification by their degree of quantification

Qualitative methods give a yes / no response, with no indication of the concentration of the putative analyte. Examples include:

- bacterial growth inhibition tests which give a result of either “no zone” or “zone of inhibition”;
- inhibition tests which give a colour change;

- immunochemical / ligand binding tests, where a response is considered as “above” or “below” a Cut-Off Level; or where analytes with different cross-reactivities are included within the method scope;
- chromatographic tests (HPLC, LC-MS/MS, ...), where a peak is considered as “present” or “absent”. They could be simply validated as qualitative screening methods as it is described in this document when quantification is not required at the screening step [6].

Semi-quantitative methods give an approximate indication of the concentration of the putative analyte. Whilst the numerical result may not be regarded as reportable, this may be useful to the analyst in deciding the calibration range for the subsequent (quantitative) confirmatory test. Examples include:

- microbial growth inhibition tests where an attempt is made to relate the size of the inhibitory zone to the putative analyte concentration;
- biochemical tests which include a calibration curve (e.g. ELISA, but only if the test is specific for a single analyte);
- chromatographic tests, calibrated over a short range which may not include the sample response;
- any physicochemical test (e.g. HPLC, LC-MS/MS, ...) where the measured method precision characteristics do not meet the requirements for quantitative tests.

Quantitative methods meet the same requirements for accuracy, dynamic range, and precision as confirmatory tests. And thus, when the quantification is required, these methods shall be validated as confirmatory methods, as detailed in the Commission Decision 2002/657/EC [1].

When the method is only used for screening purposes, the specific requirements concerning the confirmation of identity (identification points according to the section 2.3.2.2. Table 6 of the Decision EC/2002/657 [1]) are not necessarily needed.

4. Principles to be followed for the validation of screening methods

4.1. Key requirements

The key requirement for a screening method (whether qualitative or (semi-)quantitative) is its ability to reliably detect the analyte in question at the chosen Screening Target Concentration and to avoid false-compliant results. The Screening Target Concentration should be low enough to ensure that if the analyte in question is present in the sample at the Regulatory Limit, the sample will be classified as 'Screen Positive'. **Validation (whether initial or abridged) should provide the objective evidence that this key requirement is met.** Validation (both initial and abridged) must cover the entire matrix / species / analyte combinations claimed within the scope of the method SOP. **However, the extent of validation required will vary depending on whether the validation is initial or abridged.**

Minimum performance parameters for screening tests are specified in Chapter 3, Table 9 of Commission Decision 2002/657/EC [1].

As a general principle, there has to be a sufficient margin of difference between the Screening Target Concentration and the Regulatory Limit. **Therefore CCβ must be less than or equal to the Regulatory Limit.**

It is important to note that screening methods may not be able to reliably detect all relevant target analytes at the Regulatory Limit in all matrices and species. If essential analytes or species are not covered, then additional tests, using an alternative method, must be added.

4.2. Choice of analytes used for validation and selectivity of the method

The selection of analytes which will be used for either the initial or abridged validation study depends on the scope of the screening method which is described in the method SOP.

If the screening method can not distinguish between different analytes within one chemical family (e.g. tetracyclines or beta-lactams), validation should be carried out for each analyte which is considered relevant for the laboratory. For example, relevant analytes are each analyte that the laboratory might be required to include in an analytical programme for detection of residues in Control Plans.

Alternatively, validation may be performed using a number of analytes which are representative for the analyte group in question (see sections 4.2.1., 4.2.2., 4.2.3.).

4.2.1. Multi-class methods using inhibition tests

For inhibition–based multi-class methods, at least one analyte should be chosen in the validation study to represent each analyte group (e.g. for microbial growth inhibition tests, one tetracycline, one sulphonamide, one β-lactam, one aminoglycoside and one macrolide could be used). It should be noted however that in the case of microbial growth inhibition tests, not all analytes in the one antimicrobial family will display the same antimicrobial activity profile. Therefore it is recommended, prior to validation, that activity profiles are determined for all of the relevant analytes in each antimicrobial family using standard solutions at different concentrations around the MRL. These activity profiles allow at least one or two representative analytes per family of analytes to be considered for the validation study.

The analyte(s) to be selected for the validation study should ideally be the least sensitive in their class i.e. the Screening Target Concentration the closest to the Regulatory Limit. When the MRL is the same for all the family (e.g. tetracyclines), a single analyte (the least sensitive) could be chosen. For example, it has been demonstrated by CRL AFSSA-Fougères that oxytetracycline is the less sensitive tetracycline (among those having an MRL set) which can be detected by many multi-plate microbial growth inhibition tests. When different MRLs were set in one family (e.g. penicillins), several analytes should be validated and the Screening Target Concentration will be set regarding the respective MRLs. For example, in the penicillin family, ampicillin and cloxacillin were the less sensitive penicillins and both have to be chosen as representative antibiotics, with different Screening Target Concentrations.

The CRL can provide specific advice on the choice of representative analytes for bacterial growth inhibition tests [8-10].

4.2.2. Multi-class methods using biochemical tests

For biochemical tests (e.g. ELISA), which can bind several analytes with varying cross-reactivities, if all of these analytes are included in the scope of the method, initial validation must be sufficient to demonstrate that all of the analytes in question will be reliably extracted (if necessary) and detected.

Commonly encountered problems include the fact that ELISA kit manufacturers may not indicate whether the supplied information on antibody-antigen cross-reactivities has been generated in buffer (standard) solution or in biological matrix. In ELISAs with a chemical extraction step, recoveries may not be specified. Therefore it may not always be possible to calculate the detection capabilities of each of the cross-reactive analytes from the CCβ of the representative analyte.

The detection capability has to be determined for the single analyte detected by the test or for the representative analyte (e.g. the analyte with the lowest cross-reactivity). In case the test is not specific for one analyte (e.g. a multi-sulphonamides test), the cross-reactivities with different other analytes have to be determined. Finally the detection capabilities of the other analytes from the multi test could be derived from the detection capability of the representative analyte in relation with the percentages of cross-reactivities.

4.2.3. Multi-class methods using physicochemical screening techniques

Concerning physico-chemical methods which allow analytes to be differentiated on the basis of their chemical properties, the first idea is to validate for at least one analyte which should be selected from each known chemical class or sub-class (e.g. for quinolones, one acidic compound and one amphoteric compound can be chosen for the validation study).

However, by using a multi-class screening method (e.g. an LC-ToF-MS screening method), if analytes have a different retention time (Rt), they may undergo different effects (ion suppression or ion enhancement) because of different amounts of co-eluting matrix compounds. For that reason, it is advisable to test for all analytes and not for a subselection even if the analytes have very similar physicochemical properties.

4.2.4. Summary

Examples of criteria that could be used to choose the analytes included in the method that have to be validated:

- for microbial growth inhibition tests:
  - choice of the analyte(s) which give (s) the lowest inhibition in the conditions used;
  - when the method is a multiplate test, the validation study is performed at least on the most sensitive plate towards the concerned antibiotic;
- for immunological tests:
  - the analyte with the lowest cross-reactivity;
- for (semi-)quantitative methods with an extraction step:
  - the analytes with the lowest analytical recovery;
  - all the analytes included in the method when ion suppression may occur.

4.3. Preparation of "simulated tissue" for validation of microbial growth inhibition tests

For those methods with an extraction step prior to the analysis (biochemical and physicochemical methods), the matrix preparation for is rather simple with spiking samples. Even some microbial growth inhibition tests use fluid extracted from the tissue, and therefore can be validated also simply using spiked samples. However, there is a special issue with those microbial growth inhibition tests which use solid matrices (e.g. slices of whole tissue - muscle, kidney – applied directly to the plate) and where there is no extraction step. In such cases validation must be carried out using “simulated tissue”. This simulated tissue should contain the analyte at the concentration of interest and the sample should behave the same way as an incurred tissue slice, i.e. possible matrix effects could be observed and taken into account in the results.
There are two practical options for the generation of a simulated tissue.

Option 1 Tissue is minced, weighed, spiked and frozen. Pieces of frozen spiked tissue are placed directly on the plates. (NB This procedure may not be applicable to kidney samples - due to false positive results because endogenous components may be released during the mincing process - or for tests where tissue fluid is soaked onto a paper disc - due to insufficient fluid uptake from the minced sample).

Option 2 Paper discs are placed on the plates and then spiked with standard solutions. Paper discs are still wet when finally pieces of tissue are placed onto the spiked discs in a 'sandwich' format (in the following order: agar-paper disc-tissue slice). This protocol maintains the integrity of the tissue.

The CRL could provide advice on the protocol of preparation of simulated tissues for validation of microbial growth inhibition tests.

5. Validation procedure

5.1. Determination of specificity/selectivity and detection capability CCβ according to classical approach[1]

5.1.1. Number of samples required for validation

The number of "Screen Positive" Control Samples (i.e. samples spiked at the Screening Target Concentration) for each analyte depends on the degree of statistical confidence required in the result, and the relationship between the Screening Target Concentration and the Regulatory Limit. The lower the Screening Target Concentration in comparison with the Regulatory Limit, the fewer replicates are required to give the same degree of confidence that the screening test will correctly identify truly contaminated samples at the Regulatory Limit. For example:

- If the Screening Target Concentration is set at half the Regulatory/Action Limit or lower (e.g. 1/2 MRL), the occurrence of one or no false-compliant results following the analysis of at least 20 "Screen Positive" Control Samples is sufficient to demonstrate that CCβ is less than the Regulatory/Action Limit (MRL) and less than or equal to the ½ MRL;
- If the Screening Target Concentration is set between 50 % and 90 % of the Regulatory/Action Limit, at least 40 "Screen Positive" Control Samples (with no more than 2 false-compliant results) will be sufficient to demonstrate that CCβ is less than the Regulatory/Action Limit;
- If the sensitivity of the screening test is such that the Screening Target Concentration approaches the Regulatory/Action Limit (10 % below the Regulatory/Action Limit), more "Screen Positive" Control samples may be required. A maximum of 60 replicates (with no more than 3 false-compliant results) is needed to demonstrate that CCβ is fit for the purpose. These larger studies can be undertaken in sequential stages i.e. the first twenty pairs of control samples tested, and if more than one spiked sample falls below the Cut-Off level, the validation can be abandoned at this point, the Screening Target Concentration has to be increased and the validation exercise repeated.

If the screening method is applicable to one matrix but to different animal species, the 60 different samples could be taken from the different species (e.g. 20 porcine muscles, 20 bovine muscles, and 20 poultry muscles) (see section 5.1.3.).

5.1.2. Identification of the Cut-Off Level and calculation of CCβ

Validation of screening methods (whether qualitative or semi-quantitative) requires identification of a Cut-Off Level at, or above which the sample is categorised as 'screen positive' and liable to physicochemical confirmation. Two different approaches for establishing Cut-Off Levels for semi-quantitative screening tests are given in Annexes I and II.

In the case of a microbiological growth inhibition test, a typical Cut-Off Level would be an inhibition zone with a width of > 2mm. In this case, any sample giving a zone of > 2mm would be classified as 'screen positive'. All samples spiked at the screening target concentration should give zones > 2mm to be classified as 'screen positive'.

- The Screening Target Concentration (x_1) at which the matrix blank samples will be spiked in order to establish the Cut-Off Level for the analyte in question should be ideally set at **half** the Regulatory/Action Limit; if not possible, a concentration between 50 and 100 % of the Regulatory/Action Limit should be chosen.
- Select **60[2] samples** of **one matrix**. If for example the matrix is bovine muscle, each sample should result from a different batch. For the reliable determination of CCβ and specificity, at least 60 blank samples and 60 spiked samples should be analysed. Where the screening

[1] An alternative multi-factorial 'matrix comprehensive' validation model may be used. This is described in section 5.2.

[2] It is not always necessary to analyse as many as 60 samples. See Section 5.1.1.

method detects more than one analyte, this spiking exercise must be repeated for each analyte or at least for each of the analytes considered to be representative.

Step 1 Spike 60 blank samples with the analyte in question at concentration x_1;

Step 2 Analyse the 60 spiked samples and 60 blank samples according to the method SOP. These analyses should be carried out on different days and should preferably be carried out by different operators, and should ideally mimic the whole range of operating conditions likely to be encountered when using the method. It is recommended that this study is carried out in 'blind' conditions (the operators do not know which samples are spiked and which samples are blank).

Step 3

Approach 1 (see example in Annex I):

Evaluate the range of analytical responses for the blank samples and the range for the spiked samples. Select the lowest response in the spiked samples. This is the Cut-Off Level provided that the lowest response for the spiked samples does not overlap with the highest response for the blank samples. (See Annex I).

Approach 2 (see example in Annex II):

The second approach is a statistical approach which takes into account the β error of 5 %.
The analytical response Bi of the blank samples is determined for each of the investigations. Then, the mean response of the set of blanks B and the standard deviation "*SDb*" of their response are calculated. A "Threshold value" T can be calculated (see annex II).
The analytical response Yi is determined for each of the investigations of the spiked samples. Then, the mean response M and the standard deviation "*SD*" of the response of the spiked samples are calculated. A "cut-off factor" Fm can be calculated.
Positivity threshold T and cut-off factor Fm are matrix-specific.

Step 4 Identify the number of spiked samples with results *below* the Cut-Off Level. If more than 3 spiked samples out of 60 (i.e. 5%) are below the Cut-Off Level, the Screening Target Concentration chosen for the spiking study is too low as this Screening Target Concentration will not give a response above the cut off level and therefore be judged 'screen positive'.

Note: If x_1 is at the MRL (or Regulatory/Action Limit) and if more than 3 samples (out of 60) spiked at x_1 are below the Cut-Off Level, the validation study has to be abandoned for this concentration until the method has been improved.

If x_1 is half of the MRL (or half the Regulatory/Action Limit) and if more than 3 samples (out of 60) spiked at x_1 fall below the Cut-Off Level, the spiking concentration should be increased (e.g. to three quarters of the MRL) and the spiking study repeated.

Step 5 **Calculation of CCβ.** After the analysis of 60 spiked (or incurred) samples, the spiking level, (Screening Target Concentration) where ≤5% of false compliant results would be present at the Regulatory/Action Limit , is the detection capability CCβ of the method (i.e. the concentration at which there are ≤ 3 false compliant out of 60 spiked samples).

5.1.3. Determining the applicability and ruggedness of a screening method

Applicability:

In general MRLs do not differ in the same matrix type (e.g. muscle) between species but they often vary for different matrix types within the same species. Nevertheless, if CCβ has been determined for one matrix (e.g. bovine muscle) during the initial validation and the method is to be applied to the same matrix in another species (e.g. porcine muscle), an interfering matrix effect should be anticipated and it cannot be assumed that the same CCβ will apply to this new matrix. **Therefore CCβ must be established for the analyte(s) in question in this new matrix.** Again, this should be performed for each analyte the laboratory is required to include in a residue analysis programme or, at least on a selected number of analytes which are representative for the analyte group in question (see section 4.2).

Operational scheme:

Example : For the same matrix type (e.g. muscle) from four different species.

Provided the Regulatory/Action Limit is the same for all species and is the same as the original matrix, CCβ should be determined by analysing 20 blank samples (5 samples per species) and the same 20 blank samples spiked at the Screening Target Concentration used for the original matrix (5 samples per species). Then provided all blank samples are shown to be negative for the residue in question:-

- If the 20 spiked samples are all "screen positive" (i.e. exceed the Cut-Off Level) or if there is a maximum of 1 result below the Cut-Off Level, the method is applicable to the new matrices (or species), with the same CCβ as the original matrix.
- If there are 2 or more of the spiked samples which "screen negative" it can be inferred that CCβ for those species is greater than that estimated for the original matrix. In such a case the screening method should be fully validated for the new matrix (i.e. the Screening Target Concentration should be increased and the spiking study repeated).

Extension of the method to different matrix types and/or different species.

If CCβ has been determined for one matrix (e.g. bovine muscle) during the initial validation and the method is to be applied to a different matrix (e.g. liver) in either the same species or another species, there will almost certainly be a marked matrix effect and it can not be assumed that the same CCβ will apply to this new matrix. Therefore CCβ must be established for the analyte(s) in question in this new matrix. One approach to this issue is to use the matrix-comprehensive approach as described in Chapter 3.1.3 of the Annex to Commission Decision 2002/657/EC [1]. Alternatively, CCβ could be determined for each new species/matrix combination by analysing 20 blank samples (e.g. 20 porcine livers) and the same 20 blank porcine livers over-spiked at the Screening Target Concentration. This study should be carried out for each analyte or for a representative analyte of the analyte group in question. The interpretation of results is as described above. In the case the validation in the first matrix has been performed on each analyte, the matrix extension of the method could be either tested on each analyte again or reduced to a list of representative analytes (if matrix effect is not suspected). In the case the validation in the first matrix has been performed on a list of representative analytes, the same list could be used for the extension of the validation of the method to a new matrix. In both cases, the same analytes could be used for the validation in the new matrix only if these analytes are also relevant for the new matrix (e.g. one analyte relevant for the validation in bovine muscle could not be relevant for ovine muscle because the drug is not authorised to be used for the ovine species).

Ruggedness:

Ruggedness studies use the deliberate introduction of minor reasonable variations by the laboratory and the observation of their consequences on the results. Ruggedness studies should be conducted as it is recommended in the Commission Decision 2002/657/EC [1], by means of experimental plans.
Matrices or animal species could be included in the ruggedness study as factors that could influence the results. In this case, applicability study and ruggedness study are combined.
To investigate the ruggedness of a screening method, it is recommended to focus on one analyte found to be representative of the other analytes (if the method displays a wide detection spectrum).
The ruggedness should be evaluated by the analysis of at least 10 different blank materials and 10 different materials spiked (or incurred) at the level of interest. It is recommended to perform the studies for evaluating the detection capability and the specificity for this analyte as a blind test (unknown samples) at different days with different trained operators, if possible.
When it has been demonstrated that one factor has an effect on the performance of the method, the performance characteristics (specificity, detection capabilities) should be determined for this factor. Moreover, the impact of this factor on the performance characteristics has to be described in the validation report and in the final SOP.

5.1.4. Stability

When the stability of analyte(s) is known (bibliographical references or already characterised in the laboratory), there is no need to determine the stability again. Otherwise, the stability of the analyte in standard solution, and the stability of analyte in the biological matrix should be determined as detailed in the Commission Decision 2002/657/EC [1].

5.2. Determination of specificity/selectivity and detection capability CCβ according to alternative matrix-comprehensive approach.

An alternative matrix-comprehensive approach for method validation is described in Chapter 3.1.3 of the Annex to Commission Decision 2002/657/EC [1]. Using this multi-factorial approach reduces the number of runs (factor-level combinations) needed for validation [11-12].

Approach:

Firstly, eight samples are to be investigated according to the statistical design described in Chapter 3.1.3 of the Annex to Commission Decision 2002/657/EC [1].

After this initial validation study, quality assurance samples (QA-samples) are to be used for validation purposes and at least 20 QA-samples per year must be included (see section 6.3.1.). In cases where this cannot be fulfilled, the eight-sample-validation has to be repeated every year and as many QA-samples have to be investigated as possible. After one year the QA sample results can be statistically compared with the combined initial and/or transfer validation data.

An example of the validation procedure according to the alternative approach is presented in Annex III. A comprehensive description of this procedure will be available at the reference contact CRL-BVL (Berlin, Germany).

6. Transfer of screening methods between laboratories

The objective of the abridged validation is to demonstrate that the receptor laboratory is able to apply the transferred method correctly in accordance with the method SOP. The SOP used in the receptor laboratory must not deviate from the SOP developed in the originator laboratory. Concerning physico-chemical methods (e.g. LC-MS/MS) and biochemical tests (e.g. ELISA reader), the equipments could be purchased from different manufacturers. Particular attention has to be paid to the switching between instruments which could strongly influence the results (see section 2.8.).

Even though the principles of validation (determination of performance characteristics) are common to initial validation and to abridged validation of transferred methods, this chapter particularly deals with abridged validation requirements.

6.1. Conditions of transfer

When a screening method validated in one laboratory (the “originator”) is passed to a second laboratory (the “receptor”), then the receptor can use an abridged validation protocol **provided that**:

- the receptor laboratory has the **necessary equipment and skills** to use the method;
- the receptor laboratory has **full access to the original SOP and to initial validation data (e.g. validation report)**;
- the method is **used as described** in the original SOP using the **same operational conditions** (matrices, measurement technique, sample preparation, clean-up, critical equipment*), the **same Screening Target Concentrations** and, if applicable, the same Cut-Off Level(s).

*If the equipment (manufacturer, reference) is different from that of the originator laboratory, the receptor laboratory has to check if this equipment is critical or not for implementing satisfactorily the method. If the equipment is not critical, the SOP could be applied exactly as in the originator laboratory. In biochemical methods, change in equipment type and/or is apparently not critical (e.g. ELISA washer or reader). Concerning ELISA kits, the manufacturer of the kit should be the same as in

the initial validation. If the equipment is critical, the SOP could be slightly changed to reach the same performance characteristics.

Before abridged validation is undertaken in the receptor laboratory, the receptor laboratory **must** demonstrate that the operational conditions prevailing in the initial laboratory are covered by the receptor laboratory. If not, the transferred method will need to have a 'full' validation in the receptor laboratory.

The **choice of analytes** used for the abridged validation should be based on the one from the initial validation and, as a general principle, should be the worst-case examples reported by the originator laboratory. However there are some exceptions to this statement. If for example the method has been transferred to another country where different antibiotics within an antibiotic family are used more frequently compared to the originator country, it would be more appropriate to use these analytes in the abridged validation study, provided of course that these analytes are within the scope of the original SOP. The receptor laboratory may also have to validate more than one analyte, if the originator laboratory has demonstrated significant intra-class variation in response to the test. Again, in order to be eligible for abridged validation, the analytes chosen must be within the scope of the original SOP.

NB: In those cases where the receptor laboratory wishes to **modify** the initial screening method (e.g. extend the scope to **other matrices**, include other analytes etc), a full validation must be carried out in the receptor laboratory for the new matrices and analytes.

When deciding to adopt a screening method which has been developed in the originator laboratory (or when purchasing a commercially available screening test), the receptor laboratory must investigate the performance of the method by consulting scientific literature, examining data from the originator laboratory or supplier (initial validation report), and, if available, data from national and international standardisation organisations (e.g. ISO, AOAC, AFNOR).

Once the method is in place and before undertaking abridged validation, the skill of technicians and their ability to implement the method needs to be addressed. Staff must receive training on the principles of the method and how to run the test. Once trained, staff must run negative (section 2.5) and screen positive control (section 2.6) samples on several occasions to show that they are capable of running the test according to the SOP.

6.2. What is included in the abridged validation according to the classical concept

Provided that the conditions referred to in 6.1. are met, abridged validation may be carried out. By definition, abridged validation is less intensive than full (initial) validation. Its purpose is solely to indicate that the transferred method will work reliably in the receptor laboratory.

During abridged validation, only specificity and detection capabilities have to be determined on a reduced number of samples (20 samples whichever is the Screening Target Concentration) and these characteristics have to be compared to those obtained during the initial validation in the originator laboratory (see validation report or scientific articles). The transfer of the method will be valid when the performance characteristics are of the same order. A difference in the performance characteristics means that the method has not been transferred properly. In these conditions, the receptor laboratory has to take advice from the originator laboratory, until the problem is solved.

6.3. What is included in the abridged validation according to the alternative concept

Provided that the conditions referred to in 6.1. are met, abridged validation may be carried out. Principle of validation of transferred methods is the combined use of data of the initial validation and data of the abridged validation. This allows a substantial reduction of the workload for the routine laboratories. The combination of the gathered data however is only possible if there are no significant differences. If the in-house reproducibility obtained in the transfer validation is worse than 1.5 times the in-house reproducibility of the initial validation, the receptor laboratory has to reconsider its proceeding of the method.

In order to reduce the workload to the highest extent possible, a combination validation strategy can be applied. The statistical design of the abridged validation is equivalent to the design of the initial validation but takes into account only one concentration level. This means that only eight samples have to be investigated. These samples have to be fortified at or close to the MRL.

Additionally quality assurance samples (QA-samples) are to be used for validation purposes. It has to be required that at least 20 QA-samples are available per year. In cases where this cannot be fulfilled, the eight-sample-validation has to be repeated every year and as many QA-samples has to be investigated as possible.

7. Continuous verification

7.1. Quality control (QC) samples

Regardless of the type of screening test (qualitative or (semi-)quantitative), on-going quality control (QC) is vital to supplement data generated during either the initial validation or abridged validation studies. To this end, each batch of analyses should include both “Negative Control” (blank matrix) and “Screen Positive Control" Samples (spiked at the Screening Target Concentration). If the “Screen Positive Control" sample gives a “Negative” result (i.e. less than the Cut-Off Value), the batch of analyses should be discarded. Similarly, if the “Negative” control samples give a Positive result (i.e. above the Cut-Off Value), the batch of analyses should be discarded. In both cases, there should be an investigation into why the test has failed and remedial action taken. Results from these samples should be recorded continuously and these data should verify that the screening test works reliably and has a false-compliant rate of no more than 5% for the target analytes.

The choice of analytes to include in routine QC samples should follow the same rules as those selected for the initial or abridged validation exercise *i.e*. the worst-case analytes that are listed in the method scope or the most relevant analyte in a national control plan.

The use of spiked samples as QC is applicable to qualitative tests (e.g. tube tests (Premi®Test, Delvotest®, COPAN®, etc…), receptor-based tests (Tetrasensor®, Twinsensor®), semi-quantitative (e.g. ELISA kits) and quantitative tests (e.g. LC-MS/MS methods).

It is more problematic (for the reasons described in section 4.3.) to use spiked matrix QC samples for microbiological growth inhibition plate tests. In such cases, positive QC samples should at least be the spiked antibiotic standard paper discs (for each plate). However it is highly preferable to use incurred samples where possible, or spiked 'simulated tissue' slices as used during the initial and abridged validation phases.

The goal of continuous verification is to produce more than 20 quality control results within 12 months. QC samples should be stored for a period determined by the laboratory according to stability data available for the analyte/matrix. The data obtained with the QC samples should be stored and remain traceable as long as the method is used in the laboratory. This includes routine use of commercially available test kits.

The results obtained from the QC samples should be used to supplement the initial and abridged validation data. They may be used to verify:

- the method performance;
- batch quality of commercially available test kits;
- quality and stability of reagents and
- the skill and test performance of the technicians carrying out the analyses.

Within 12 months of routine use of the method, all available validation data (from initial validation, abridged validation and quality control) should be statistically analysed. If the number of quality control samples is below 20, additional validation samples must be carried out. The sum of initial validation results, abridged validation results and ongoing QC results for the "Screen Positive Control" samples should raise at least 40 samples during the first year. Then the number of QC should reach 20 samples each subsequent year. If the test is working reliably and is robust, no more than 5% of these 40 or 20 "Screen Positive Control" samples should fall below the Cut-Off Level of the test. If the

method is not currently employed in the laboratory, the number of QC could be reduced during the subsequent years.

7.2. Proficiency tests

Regular participation in relevant proficiency tests is strongly recommended. On an annual basis the CRLs for analytes run proficiency tests for selected analytes for which they have responsibility. There are also other commercial providers in the veterinary drug analyte field and the relevant CRLs can give further information on the options available. Investigations and corrective actions should be undertaken and documented whenever questionable or unsatisfactory results are obtained.

8. Validation report

When a screening method has been validated, either in the originator laboratory (initial validation) or in the receptor laboratory (abridged validation), a report of the validation study should be drawn up. The initial validation report should:

- identify the application range of the method, including the ruggedness statements, concentration range, matrices, species, matrix conditions and laboratory conditions;
- describe the validation study design, including the prerequisites, assumptions and formulae used in the design of an experimental plan;
- provide and summarise the results for all validation parameters,
- identify conditions which do not allow reliable analysis to be performed;
- address interferences observed during validation studies or during the analysis of quality control samples (these ongoing QC data will be added to the validation report later);
- establish for inhibitor tests a list of the different analytes which have produced a result above the cut-off level for each analyte/matrix combination on each plate;
- if applicable, provide results of participation in proficiency tests.

When a screening method has been transferred, the receptor laboratory should have access to the original validation report. Additionally the receptor laboratory should:

- document the source of the method (i.e. from where it was transferred);
- cross reference the initial validation data, with data used from scientific publications or other sources (including references);
- document on the results of the abridged validation study which it has carried out to verify that the method works reliably in their hands.

9. References

[1] EC 2002. Commission Decision 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC. *Official Journal of the European Community.* L 221:8-36.

[2] EC 2009. Council Regulation (EC) No 470/2009 of 6 May 2009 laying down Community procedures for the establishment of residue limits of pharmacologically active subtances in foodstuffs of animal origin, repealing Council Regulation (EEC) No 2377/90. *Official Journal of the European Community.* L 152:11-22.

[3] EEC 1990. Council Regulation (EEC) N° 2377/90 of 26 June 1990: laying down a Community procedure for the establishment of maximum residue limits of veterinary medicinal products in foodstuffs of animal origin. *Official Journal of European Community.* L224:1-8.

[4] EC 2005. Commission Decision 2005/34/EC of 11 January 2005 laying down harmonised standards for the testing for certain residues in products of animal origin imported from third countries. *Official Journal of the European Community.* L 16: 61-63.

[5] CRL Guidance Paper of 7th December 2007. CRLs view on state of the art analytical methods for national residuecontrol plans, 1-8.

[6] M. Gaugain-Juhel, B. Delépine, S. Gautier, M.P. Fourmond, V. Gaudin, D. Hurtaud-Pessel, E. Verdon, P. Sanders. 2009. Validation of a liquid chromatographic-tandem mass spectrometry screening method to monitor 58 antibiotics in milk : qualitative approach. Food Additives & Contaminants: Part A. *In press.*

[7] V. Gaudin, C. Hedou, and P. Sanders. 2007. Validation of a Biacore Method for Screening Eight Sulfonamides in Milk and Porcine Muscle Tissues According to European Decision 2002/657/EC. JOURNAL OF AOAC International. 90 (6) 1706-1715.

[8] V. Gaudin, P. Maris, R. Fuselier, J.L. Ribouchon, N. Cadieu, A. Rault. 2004. Validation of a microbiological method : the STAR protocol, a Five Plate Test, for the screening of antibiotic residues in milk Food Additives & Contaminants, 21 422-433.

[9] V. Gaudin; C. Hedou; A. Rault; P. Sanders; E. Verdon. 2009. Comparative study of 3 screening tests, Explorer® test, Premi®Test 2 microbiological tube tests and a multi-sulphonamides ELISA kit, for the detection of antimicrobial and sulphonamide residues in eggs. Food Additives & Contaminants: Part A. 26 (4) 427–440.

[10] V. Gaudin, C. Hedou and E. Verdon. 2009. Validation of a wide-spectrum microbiological tube test, the EXPLORER® test, for the detection of antimicrobials in muscle from different animal species. Food Additives & Contaminants: Part A.26 (8) 1162–1171. .

[11] B. Jülicher, P. Gowik and S. Uhlig. 1998. Assessment of detection methods in trace analysis by means of a statistically based in-house validation concept. Analyst, 120, 173.

[12] P. Gowik, B. Jülicher and S. Uhlig. 1998. Multi-residue method for non-steroidal anti-inflammatory drugs in plasma using high performance liquid chromatography-photodiode-array detection. Method description and comprehensive in-house validation. J. Chromatogr., 716, 221.

10. Annexes

Annex I. Determining Cut-Off Levels and CCβ in a semi-quantitative screening test.

Example A:

- MRL = 1.0 µg/kg
- Desired Screening Target Concentration = 0.5 µg/kg

Twenty (or multiples thereof) different blank matrix samples are selected. Replicates of these samples are spiked at the Screening Target Concentration, in this case 0.5 µg/kg.

Sample Number	Negative Samples	Spike @ 0.5 µg/kg
1	0.000	0.355
2	0.090	0.252
3	0.000	0.532
4	0.000	0.554
5	0.000	0.408
6	0.070	0.501
7	0.000	0.524
8	0.015	0.559
9	0.000	0.471
10	0.010	0.661
11	0.070	0.642
12	0.129	0.724
13	0.046	0.596
14	0.034	0.599
15	0.041	0.640
16	0.137	0.750
17	0.112	0.655
18	0.120	0.660
19	0.132	0.695
20	0.063	0.635

The matrix blank samples and spiked samples are analysed, preferably over a number of different days. The range of responses in the blank samples is examined. The **highest response in the blank** samples is noted – in this case it is 0.137 units. The **lowest response in the spiked** samples is noted – in this case it is 0.252 units.

In the case shown none of the responses of the spiked samples overlaps with the range of responses of the blanks. Therefore we can say that **the CCβ of this screening method is less than or equal to 0.5 µg/kg**.

In the example shown we can see that the lowest response is 0.252. Therefore **the Cut-Off Level of this test is 0.252 units.** Any sample giving a response greater than this level is deemed to be a 'screen positive' and exceeds the CCβ of the screening method.

For this test, as a batch acceptability criterion, the response generated by the “Screen Positive Control Sample" must be ≥ 0.25 units otherwise the batch is rejected.

Example B:

- MRL = 1.0 µg/kg
- Desired Screening Target Concentration = 0.5 µg/kg

In this example the **highest response in the blank** samples is 0.137 units. However the **lowest response in the spiked** samples is noted – in this case it is 0.132 units.

Sample Number	Negative Samples	Spike @ 0.5 µg/kg
1	0.000	0.355
2	0.090	0.132
3	0.000	0.532
4	0.000	0.554
5	0.000	0.135
6	0.070	0.501
7	0.000	0.524
8	0.015	0.559
9	0.000	0.471
10	0.010	0.661
11	0.070	0.642
12	0.129	0.724
13	0.046	0.596
14	0.034	0.599
15	0.041	0.640
16	0.137	0.750
17	0.112	0.655
18	0.120	0.660
19	0.132	0.695
20	0.063	0.635

In this case there is an overlap between the two sample populations which is greater than 5% (the responses of two of the spiked samples are less than the highest response in the blank samples).

A clear Cut-Off Level can not be established (due to the overlap of responses between blank and spiked samples). From these data it can be inferred that **CCβ must be greater than 0.5 µg/kg** and the **Screening Target Concentration of 0.5 µg/kg can not be reliably detected using this method**.

Either the method must be modified, or the validation study repeated using a higher Screening Target Concentration (provided this can be kept at or below the MRL).

Annex II. Determining Cut-Off Levels and CCβ in a semi-quantitative screening test.

Threshold value T:

$T = B + 1.64 \times SDb$ or technical threshold.

B the mean response and "SDb" the standard deviation of blank samples.

Cut-off factor Fm:

$Fm = M - 1.64 \times SD$

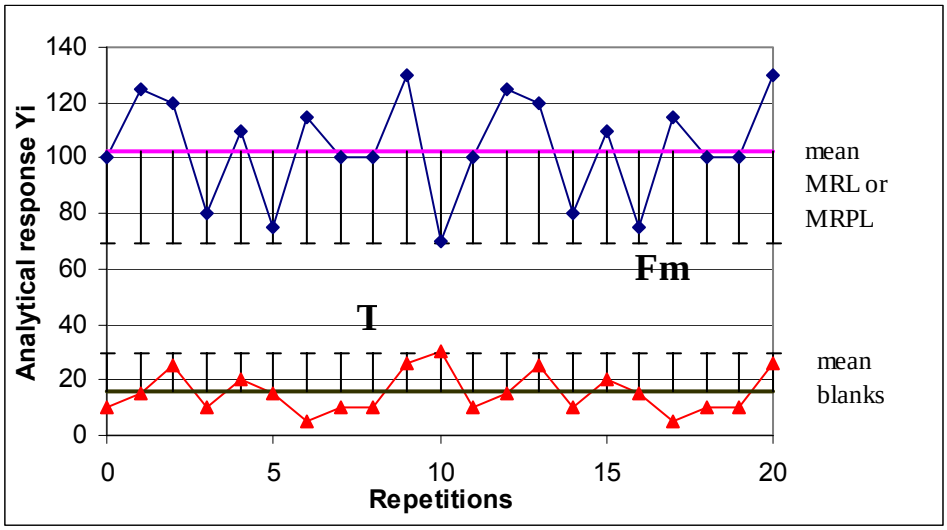
M the mean response and "SD" the standard deviation of spiked samples.

For ELISA tests, the response (B/B0 %) is inversely proportional to the concentration. Therefore:

$Fm = M + 1.64 \times SD$

Threshold value T and cut-off factor Fm are matrix-specific.

Graphical representation of threshold value T and "cut-off" factor Fm.



Between the mean of blanks B and T the false positive rate is higher than 5 %.

According to the Commission Decision 2002/657/EC [1], the detection capability is validated when: Fm > B.

Also the laboratory has to determine the rate of false positive (FP) which is acceptable with the method.
If B < Fm < T, the false positive rate is higher than 5 %.
In case of Fm > T, the rate of FP is below 5 %.

Annex III. Validation of quantitative and semi-quantitative methods according to the alternative approach.

Validation according to the alternative approach is derived from the principles of the experimental design used for confirmatory methods (cf. alternative approach in Commission Decision 657/2002/EC).

According to this approach, at least 8 samples have to be selected following an orthogonal design. Each sample has to be divided into at least 4 aliquots and to be spiked on 4 concentrations around the MRL. If samples are no blank samples, their content has to be tested by means of a reference method.
It is recommended to include non-spiked blank matrix samples in order to get a measure of the method behaviour regarding blank samples. This serves to determine the false-positive rate and thus also serves economic ends, i.e. for each sample in the experimental plan, 5 aliquots have to be analysed. The total number of analyses then increases to 40.

The 32 or 40 sub-samples are processed on different days according to the orthogonal design. In the framework of this experimental plan, matrix and/or species can be varied on 2 levels each, forming 2 factors. Moreover the different conditions in the laboratories in which the method is to be used have to be taken into account by laying down further factors, each with two factor levels. When selecting the factors, exclusively noise factors (which cannot be controlled in routine analysis) are to be taken into consideration. This includes for example fluctuations in temperature or the operators' different skills.

It should be noted that skill and matrix influence the measured response through many different “paths”. Quite often they provoke interactions with specific measurement conditions (e.g. certain skills are required only for certain matrices) or a change of precision under repeatability or reproducibility conditions. Therefore when it comes to design and analysis, one has to be aware that results might be affected not only by simple factorial “main” effects as in a ruggedness study, but also by interaction effects or “dispersion” effects (change of precision).

Up to 7 factors with principally predictable (reproducible) effects (such as incubation temperature or incubation time or skill) can be taken into account if the study comprises 8 different samples. With more samples, more factors with predictable effects can be taken into account. However, 3 to 7 factors are sufficient if they represent the major error sources.

In addition, effects from factors with unpredictable effects (e.g. effects from different technicians with equal skills or different lots or different measurement days) have to be taken into account. Factors with unpredictable effects (e.g. different lots of reagents or media) should be varied 8 times.[3] Only the combined effect of these factors can be quantified.

With these factors and the corresponding variation levels the experimental plan is designed.

The following table describes a typical experimental design for 6 factors with predictable effects and 2 factors with unpredictable effects. Factor levels 1 and 2, respectively, represent the two categories or the two levels (high-low) of the respective factor.

Factors with predictable effects						Factors with unpredictable effects			Spike levels [ppm]				
Species	Matrix	Storage of sample	Tempe-rature	Incuba-tion time	Skill	Day	Lot of reagent A	Lot of reagent B					
1	1	1	1	1	1	1	1	1	0	10	20	30	40
2	2	2	1	1	1	2	2	2	0	10	20	30	40
1	2	1	2	1	2	3	3	3	0	10	20	30	40
1	2	2	2	2	1	4	4	4	0	10	20	30	40
2	2	1	1	2	2	5	5	5	0	10	20	30	40
1	1	2	1	2	2	6	6	6	0	10	20	30	40
2	1	2	2	1	2	7	7	7	0	10	20	30	40
2	1	1	2	2	1	8	8	8	0	10	20	30	40

[3] In case that not 8 lots are available, one might consider to deviate from the prescribed design and to use 4 lots only.

A statistical evaluation of the data is performed in order to provide both the required parameters and the concentration-response curve for the determination of the false non-compliant rate at different concentrations.
The evaluation is based on a generalised mixed linear model which is described in [Gowik/Uhlig 2009]. The calculations can be carried out by means of a special software, e.g. *InterVal bioscreen.*

If the outcome of the evaluation is not satisfactory with regard to the rate of false non-compliant or to the rate of false compliant, another cut-off level may be established. Then the calculation of the false-compliant rate and of the false non-compliant rate has to be repeated.

Note:

The reference contact for detailed information about the alternative matrix-comprehensive approach is the CRL Berlin.

References:

Uhlig/Gowik 2009: Factorial Interlaboratory Tests for Microbiological Measurement Methods Journal of Consumer Protection and Food Safety, 2010, DOI 10.1007/s00003-009-0524-z

InterVal bioscreen 2009, Software for in-house validation of biochemical screening methods. quo data GmbH, www.quodata.de.