EURL Guidance Document on Screening Method Validation

1 Scope ... 3
2 Definitions ... 3
Screening Target Concentration ... 3
2.1 3
2.2 Detection capability for screening (CCβ) ... 3
2.3 Threshold level T ... 4
2.4 Cut-off level Fm ... 4
2.5 Decision value (DV) ... 4
2.6 “Screen negative control" sample (blank sample) ... 4
2.7 “Screen positive control" sample ... 4
3 Screening method classification and key performance characteristics ... 4
4 General considerations for validation ... 8
4.1 Key requirements ... 9
4.2 Choice of analytes used for validation of the method ... 11
4.2.1 Microbial growth inhibition tests ... 11
4.2.2 Biochemical tests (ie. immunological) ... 11
4.2.3 Multi-class physico-chemical methods ... 12
5 Validation without calibration curve (Cases I and II) ... 12
5.1 Principle ... 12
5.2 Samples and experiments ... 12
5.3 Identification of Cut-Off level Fm and determination of CCβ ... 13
5.3.1 Method 2 ... 13
5.4 Other validation parameters ... 14
5.4.1 Specificity/selectivity ... 14
5.4.1.1 False positive rate ... 14
5.4.1.2 Cross-reactions (CR) for biochemical tests ... 15
5.4.2 Method applicability ... 16
5.4.3 Stability ... 17
5.4.4 Ruggedness ... 17
6 Validation with calibration curves (Cases III and IV) ... 17
6.1 Principle ... 17
6.2 Samples and Experiments ... 17
6.3 Determination of CCβ ... 18
6.3.1 Method 1 ... 18

6.3.2 Method 2 ... 18
6.3.3 Method 3 ... 19
6.4 Other validation parameters ... 20
6.4.1 Specificity/selectivity ... 20
6.4.2 Trueness, precision, relative matrix effect, absolute recovery ... 20
6.4.3 Stability ... 20
6.4.4 Ruggedness ... 20
7 Fitness for purpose ... 20
8 Validation report ... 21
9 Ongoing method performance verification ... 21
10 Extension of methods ... 22
11 References ... 24
12 Annexes ... 26
12.1 Validation of screening methods without calibration curves (cases I and II (Table 1)) ... 26
12.1.1 Validation of qualitative and (semi-)quantitative methods with one validation standard at STC (Method 2) ... 26
12.1.2 Validation of qualitative and (semi-)quantitative methods with several validation standards as potential STC (Method 2) ... 27
12.1.3 Validation of (semi-)quantitative methods by using the response information in a semi-quantitative way for the calculation of threshold level T, cut-off level Fm (Method 2) ... 28
12.2 Validation of screening methods with calibration curves (Cases III and IV (Table 1)) ... 31
12.2.1 Example of experimental plan for semi-quantitative and quantitative methods with calibration curves (cases III and IV) in accordance with the classical validation approach implementing Method 1 or 3 ... 31
12.2.2 Example of a validation of quantitative and semi-quantitative methods (cases III and IV) implementing Method 1 or 3 in accordance with the factorial validation approach ... 32

1 Scope

This guideline document supplements Commission Implementing Regulation 2021/808 [1] regarding the validation of screening methods. This guideline covers the validation of residue analysis by three types of screening methods: Qualitative, semi-quantitative and quantitative screening methods. The techniques which can be applied are divided in biological, biochemical and physico-chemical methods.

The objectives of this guideline document are to define:

- the minimum requirements to be fulfilled by the full validation;

This guideline document includes:

- a 'full validation' protocol for demonstrating performance characteristics for newly developed or introduced screening methods;
- recommendations on routine quality control (continuous verification) for screening methods.

2 Definitions

2.1 Screening Target Concentration

'Screening Target Concentration'(STC) means the concentration lower than or equal to the CCβ at which a screening measurement categorises the sample as potentially non-compliant 'Screen Positive' and triggers a confirmatory testing [1].

1- For authorised analytes, the STC shall be lower than or equal to the level of interest (LoI) (defined in CIR (EU) 2018/470 [2]) (e. g. MRL, ML) (and should preferably be set at one half of the level of interest wherever possible or as low as possible).
2- For prohibited & unauthorised analytes, the STC shall be as low as reasonably achievable and in any case lower than or equal to the RPA or MMPR (if available).
3- For authorised substances, for which no MRL has been set in a specific matrix/species, the MMPR is 1/4th of the cascade MRL (established under CIR (EU) 2018/470 [2]) is the relevant level. The STC shall be lower than or equal to the level of interest (with the target 0,1 times cascade MRL, where reasonably feasible) [3].
4- For analytes for which MRLs have not been established according to Council Regulation (EC) No 470/2009 [4], the STC should be as low as reasonably achievable.

The further the STC is below the level of interest, the lower the probability of obtaining a false compliant (i.e. false-negative) result in samples containing the drug at the Regulatory Limit.
A maximum STC (STCmax) can be calculated which still fulfils the requirement for a screening method (CCβ ≤ MRL/RPA/MMPR).

2.2 Detection capability for screening (CCβ)

'Detection capability for screening (CCβ)'means the smallest content of the analyte that may be detected or quantified in a sample with an error probability of β [1]:

(a) in the case of prohibited or unauthorised pharmacologically active substances, the CCβ is the lowest concentration at which a method is able to detect or quantify, with a statistical certainty of 1 – β, samples containing residues of prohibited or unauthorised substances;

(b) in the case of authorised substances, the CCβ is the concentration at which the method is able to detect concentrations below the permitted limit with a statistical certainty of 1 – β;

2.3 Threshold level T

The threshold level T is calculated from the response, the signal or a concentration of replicate blank samples. The threshold level T must be different from the cut-off level Fm.

2.4 Cut-off level Fm

The cut-off level Fm is the response, the signal or a concentration from a screening test which indicates that a sample contains an analyte at or above the STC. If the Cut-Off Level and hence the corresponding STC is exceeded a subsequent confirmatory test is carried out. During the full validation process, the cut-off level Fm may be established through analysis of matrix blank samples and replicates of those same samples spiked (fortified) at the STC.
The established cut-off level Fm is calculated so that the probability for a response (signal or concentration) below the cut-off level Fm for analysis of samples truly containing analytes at the STC is not larger than 5 % (i.e. in this case the beta error criterion for screening methods is fulfilled and STC = CCβ).

2.5 Decision value (DV)

The decision value is a response, a signal or a concentration which corresponds to a true or fortified concentration and the linkage of which is represented by a calibration curve.
The decision value for a screening test is calculated for a given STC via the calibration curve.

2.6 “Screen negative control" sample (blank sample)

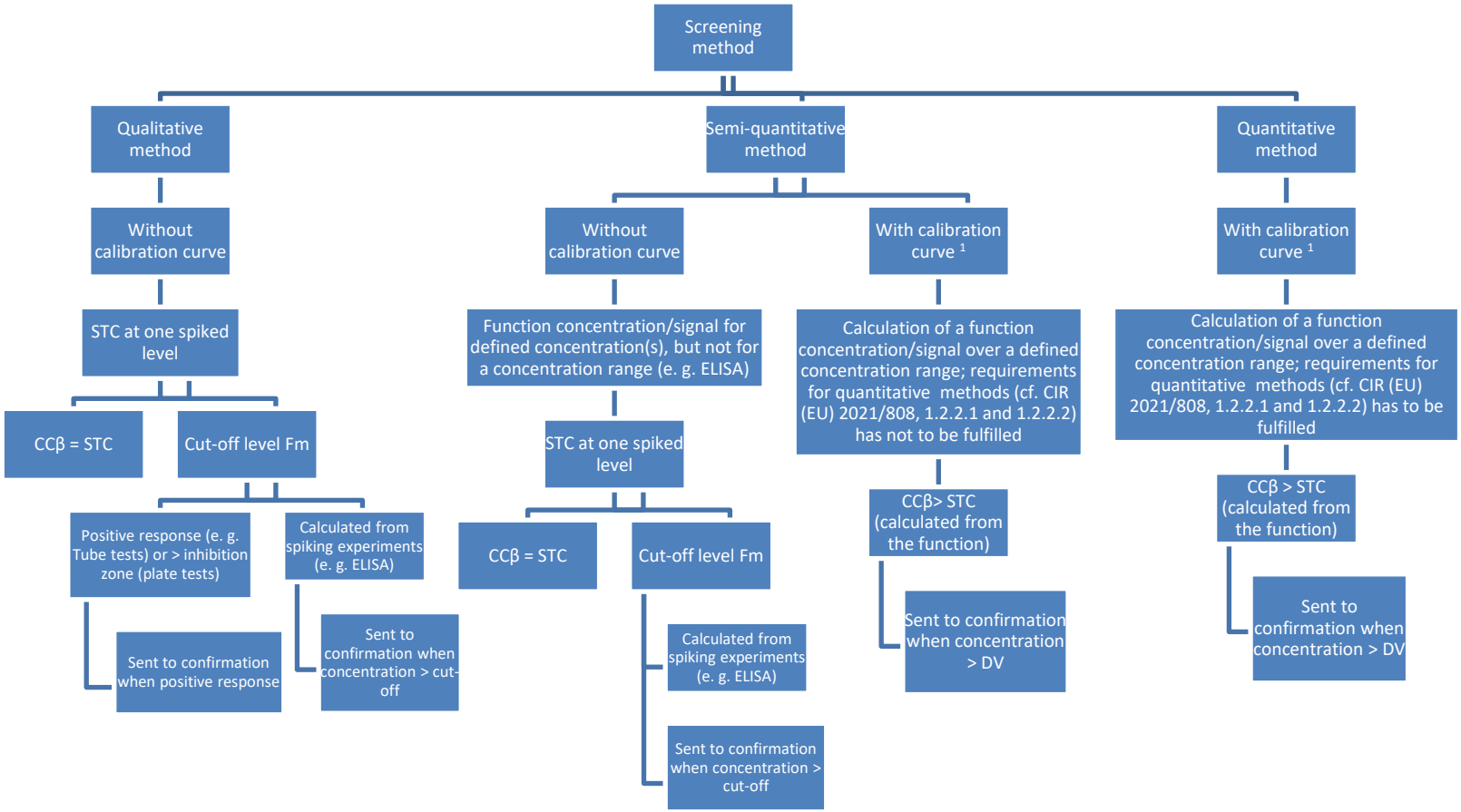
These are samples from animals of known history which have not been exposed to the substance in question. If samples from such animals are not available, samples which have been previously confirmed as compliant and not containing residues of the substance of interest by suitably sensitive physico-chemical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

2.7 “Screen positive control" sample

These are blank quality control samples which are fortified at a concentration at or above the STC and ideally at the CCβ [1]. 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 (if concentration ≥ STC). When screen positive control samples are run in the screening test, they should be classified as 'screen positive' (suspicious) if the test is operating correctly.

3 Screening method classification and key performance characteristics

Chapter 1 of Annex I to CIR (EU) 2021/808 [1] describes the requirements for analytical methods for the determination of pharmacologically active substances in foods of animal origin. Table 5 of this regulation gives a categorisation of screening methods according to their degree of quantification.
The key performance characteristic for the evaluation of a screening method is the detection capability CCβ. Three different approaches are proposed in CIR (EU) 2021/808 (chapter 2.7. Detection capability for screening (CCβ)) [1] for the calculation of CCβ: **methods 1, 2 and 3**.
The decision tree (**Figure 1**) allows to select the appropriate method for the planned validation study with a reference to the related chapters of this guidance. The following table (**Table 1**) summarises the meaning of CCβ, STC and cut-off-level Fm depending on the selected screening method category.



1

Figure 1. Decision tree for the selection of one approach to the determination of detection capability CCβ.

[1] Method 2 of CIR (EU) 2021/808 (2.6.) for the determination of detection capability CCβ is also applicable to semi-quantitative and quantitative methods with calibration curves; considering one of the calibrants generally the LCL (lowest calibration level) as the relevant STC for a screening purpose (Chapter 6.3.2. of this guidance). In this case, CCβ = STC if the criteria for determination of CCβ are fulfilled.

Table 1. Four different cases for the validation of screening methods in relation to their degree of quantification. Note that (Semi-)quantitative methods (cases III and IV) can also be evaluated as qualitative methods which implies that the calculation of CCβ and the analytical result from which a sample has to be confirmed is to be done according to case I or II as given in the table.

	Screening Method	Calibration curve for quantification	Cut Off Level	Concentrations for validation (STC)	CC_β	Sent to Confirmation	Remarks
I	**Qualitative Methods**	without	Positive response (e. g. Tube tests) or > inhibition zone (plate tests)	One spike level STC	=STC	« positive » (e. g. Tube tests) or > inhibition zone (plate tests)	Binary response (+ or -, 0 or 1)
II	**Semiquantitative Methods**	without	Cut Off Level Fm calculated from spiking experiment	One spike level STC	=STC	« positive » (≥ Cut-Off Level Fm)	Function concentration/signal for defined concentration(s), but not for a concentration range (e. g. ELISA)
III	**Semiquantitative Methods** [2]	with	Decision value (DV)	Several options : a range of concentrations around STC	> STC (calculated from the function)[1]	≥ DV	Calculation of a function concentration/signal over a defined concentration range; **requirements for quantitative methods (cf. CIR (EU) 2021/808, 1.2.2.1 and 1.2.2.2 (x))**

[2] Method 2 of CIR (EU) 2021/808 (2.6.) for the determination of detection capability CCβ is also applicable to semi-quantitative and quantitative methods with calibration curves; considering one of the calibrants generally the LCL (lowest calibration level) as the relevant STC for a screening purpose (Chapter 6.3.2. of this guidance). In this case, CCβ = STC if the criteria for determination of CCβ are fulfilled.

							do not have to be fulfilled
IV	**Quantitative Methods** [3]	with	Decision value (DV)	Several options : a range of concentrations around STC	> STC (calculated from the function)[2]	≥ DV	Calculation of a function concentration/signal over a defined concentration range **requirements for quantitative methods (cf. CIR (EU) 2021/808, 1.2.2.1 and 1.2.2.2) have to be fulfilled**

Method 2 of CIR (EU) 2021/808 (2.6.) for the determination of detection capability CCβ is also applicable to semi-quantitative and quantitative methods with calibration curves; considering one of the calibrants generally the LCL (lowest calibration level) as the relevant STC for a screening purpose (Chapter 6.3.2. of this guidance). In this case, CCβ = STC if the criteria for determination of CCβ are fulfilled.

(Semi-)quantitative methods can also be validated as qualitative screening methods as it is described in this document (chapter 5. Validation without calibration curves) when quantification is not required at the screening step [5]. The detection capability CCβ can in this case be evaluated using "method 2" (cf 2.7 CIR (EU) 2021/808 [1], 2.7 – here referenced in chapter 6.3.2.). It should be noted that if it is decided to validate a (semi-)quantitative method in a qualitative manner, the (semi-)quantitative results of this method also have to be disregarded after validation when the method is applied in routine.

For semi-quantitative methods, 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.

In the following examples of methods based on their degree of quantification are given.

Table 2. Classification of screening methods by their degree of quantification.

Degree of quantification	**Characteristics**	**Examples**
Qualitative yes / no response	No indication of the concentration of the putative analyte	- 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/biosensors - any physico-chemical method (e. g. HPLC, LC-MS/MS, ...) including chromatographic tests where a peak is considered as "present" or "absent"
Semi-quantitative gives an approximate indication of the concentration of the putative analyte	Without calibration curve With calibration curve	- biochemical tests, used with or without a standard curve (e. g. ELISA, but only if the test is specific for a single analyte); - any physico-chemical method (e. g. HPLC, LC-MS/MS, ...) including chromatographic tests, calibrated over a short range which might not include the sample response;
Quantitative gives a concentration of the putative analyte	With calibration curve	- Physico-chemical methods (e. g. HPLC, LC-MS/MS, ...)

4 General considerations for validation

Chapter 2 of Annex I to CIR (EU) 2021/808 [1] describes the performance characteristics to be determined for the validation of analytical methods for the determination of pharmacologically active substances in foods of animal origin. Minimum performance characteristics to be determined for screening methods are specified according to the classification of the method with regard to the degree of quantification.
For all screening methods, selectivity/specificity, the detection capability CCβ, stability and ruggedness have to be determined. Trueness, precision, relative matrix effect/ absolute recovery need to be determined for quantitative screening methods. Precision shall also be determined for semi-quantitative screening methods to prove the suitability of the method for avoiding false compliant analytical results, but the precision requirements of Chapter 1.2.2.2 of the CIR (EU) 2021/808 [1] do not need to be met.
For confirmatory methods used as screening methods, the same performance characteristics as other screening methods have to be determined and especially detection capability CCβ which is not

determined for confirmatory methods. A strategic planning of the validation experiments required for the confirmation method validation usually allows to validate the method in terms of screening simultaneously without the need for additional experiments. If calibrations curves with confidence intervals are available (issued from the validation of the confirmatory method), the CCβ can easily be calculated for given STCs with methods 1 and 3 (cf 2.7 CIR (EU) 2021/808 [1]).

In contrast to confirmatory analysis, there are no restrictions for screening methods with regard to the suitability of specific approaches or detection principles. Table 3 gives an overview on possible screening methods differentiated by the detection principle.

Table 3. Classification of screening methods by their detection principle.

	Principle	**Characteristics**	**Examples**
Biological methods	Detect cellular responses to analytes (e. g. oestrogenic effect, inhibition of bacterial growth, cellular effect, hormonal effect)	Not selective and can cover several chemical classes of active analytes (e. g. hormones, antimicrobials). They do not allow the identification of individual analytes.	Inhibition tests: Plate tests, tube tests, microplate tests
Biochemical methods	Detect molecular interactions (e. g. antigens, proteins) between analytes and antibodies or receptor proteins. Chemical labelling of either the analyte or antibody/receptor allows the interaction to be monitored and measured.	Either selective for one analyte, a group of analytes exhibiting similar molecular structures or analyte-specific	ELISA, radioimmunoassays (RIA), biosensors, FCIA (flow cytometry immunoassay)
Physico-chemical methods	Distinguish the molecular structure of analytes by detection of signals related to molecular characteristics	Able to distinguish between similar molecular structures and allow the simultaneous analysis of several analytes (e. g. multi-residue methods).	GC, HPLC-UV-DAD, HPLC- FLD, $MS^{(n)}$, trap MS, ToF MS, other hybrid MS, TLC.

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 or below the level of interest with a false-compliant results rate ≤ 5 % (i.e. a maximum β error of 5 %).
Validation should provide the objective evidence that this key requirement is met with the selected STC. Validation must cover all matrix / species / analyte combinations claimed to fall within the scope of the method.

An overview on all required performance characteristics is given in Table 4 (cf. Table 5 in CIR (EU) 2021/808 [1]).

Table 4. Performance characteristics and respective acceptance criteria.

Performance characteristic	**Acceptance criteria**
Detection capability CCβ	Qualitative criteria - authorised substances: lower than or equal to the MRL / ML - prohibited / unauthorised substances with RPA: as low as reasonably achievable and in any case lower than or equal to the RPA - authorised substances, for which no MRL has been set in a specific matrix/species: : lower than or equal to the MMPR (cf. MMPR guidance [3]) - prohibited / unauthorised substances without RPA: as low as reasonably achievable (cf. MMPR guidance [3])
Specificity / selectivity	No fixed criteria. The results for the parameter shall be evaluated using expert knowledge. The responsible scientist shall identify critical aspects which may require method improvements.
Trueness‡	Concentration dependant, see 1.2.2.1, Annex 1 of CIR (EU) 2021/808
Precision‡ (x)	Concentration dependant, see 1.2.2.2, Annex 1 of CIR (EU) 2021/808
Relative matrix effect‡*	See 2.10, Annex 1 of CIR (EU) 2021/808
Absolute recovery‡*	No fixed criteria. The results for the parameter shall be evaluated using expert knowledge. The responsible scientist shall identify critical aspects which may require method improvements.
Stability	See 2.5, Annex 1 of Commission Implementing Regulation (EU) 2021/808, experiments for stability are not included in the exemplary validation plans given in this guidance.
Ruggedness	No fixed criteria. The results for the parameter shall be evaluated using expert knowledge. The responsible scientist shall identify critical aspects which may require method improvements.

‡ Only relevant for quantitative screening methods.

(x) The precision requirements of Chapter 1.2.2.2 of CIR (EU) 2021/808 do not need to be met for semi-quantitative screening methods. However, the precision shall be determined to prove the suitability of the method for avoiding false compliant analytical results.

*Relevant for quantitative screening methods to prove by means of the validation that the requirements for the performance characteristics are met. The relative matrix effect of the method shall be determined when this effect was not assessed during the validation procedure. The absolute recovery of the method shall be determined when no internal standard or no matrix-fortified calibration is used.

The CCβ shall be determined for screening methods. Several approaches arising from four different cases in relation to the degree of quantification of the screening method have to be considered (cf Table 1). For the calculation of detection capabilities CCβ, CIR (EU) 2021/808 (2.7. Detection capability for screening (CCβ)) suggests three different options for authorised substances and for unauthorised and prohibited substances [1]. Chapter 5 describes the general approach for the determination of the CCβ and related quantities in line with **Method 2** (applicable to cases I and II of Table 1) given in CIR (EU) 2021/808 [1], chapter 6 the determination of the CCβ in line with **Methods 1, 2 and 3** (applicable to cases III and IV of Table 1).

4.2 Choice of analytes used for validation of the method

If the screening method cannot 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.

Table 5. Choice of analytes in relation to the principle of the screening method.

Type of screening method	Analytes to be validated
Microbial growth inhibition tests	The analyte(s) which give (s) the lowest inhibition in the conditions used; when the method is a multi-plate test, the validation study is performed at least on the most sensitive plate towards the concerned antibiotic
Biochemical tests (ie. Immunological)	The analyte with the lowest cross-reactivity
Physico-chemical methods	All relevant analytes

4.2.1 Microbial growth 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, one quinolone 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 that activity profiles are determined for all of the relevant analytes in each antimicrobial family using standard solutions at different concentrations around the MRL prior to validation. 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. their STC is the closest to the regulatory limit. When the MRL is the same for the whole substance class (e. g. sulfonamides), a single analyte (the least sensitive) could be chosen. When different MRLs have been established within the same substance class (e. g. penicillins), several analytes should be validated and the STC will be set regarding the respective MRLs.
The EURL can provide specific advice on the choice of representative analytes for bacterial growth inhibition tests [6-9].

4.2.2 Biochemical tests (ie. immunological)

For biochemical tests (e. g. ELISA, biosensors), 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 sufficiently demonstrate that **all of the analytes in question** will be reliably extracted (if necessary) and detected.
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 physico-chemical methods

As analytical methods based on physico-chemical principles (e. g. methods based on chromatography and/or mass spectrometry) are able to identify certain analytes, all analytes which are to be included in such a method must also be included in the validation. It is applicable to qualitative, semi-quantitative and quantitative physico-chemical methods.

Two examples of validation procedures are proposed in this validation guideline as it is already recommended in CIR 2021/808 [1]: the conventional validation (2.2.1. in CIR 2021/808) and the validation according to alternative models (2.2.2. in CIR 2021/808). Either approach can be used to validate a screening method. Other approaches to demonstrate that the method complies with performance criteria may be used, provided that they achieve the same level and quality of information.

5 Validation without calibration curve (Cases I and II)

5.1 Principle

For qualitative and semi-quantitative methods without calibration curves, the main performance characteristics to be determined are detection capability CCβ and specificity/selectivity.
The calculation of the performance characteristics in accordance with the requirements listed in Table 4 requires the performance of several individual experiments. Therefore, in order to minimise the workload, it is advised to combine experiments as much as possible (e. g., detection capability CCβ and testing for specificity).

5.2 Samples and experiments

For the validation experiments, a number of different blank and spiked samples is required, which are prepared according to the method description of the analytical method to be validated.
For those microbial growth inhibition tests which use solid matrices (e. g. slices of whole tissue - muscle, kidney – applied directly to the plate) without an extraction step prior to the analysis, validation must be carried out using “simulated tissue”. This simulated tissue should contain the analyte at the concentration of interest (STC) 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.
The EURL Fougères recommends the following protocol for the generation of simulated tissues: tissue is minced, weighed, spiked and frozen. Pieces of frozen spiked tissue are placed directly on the plates. (note that this procedure may not be applicable to kidney samples - due to false positive results triggered by endogenous components 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).
The STC at which the matrix blank samples will be spiked in order to establish the cut-off level Fm for the analyte in question should ideally be set at half the LoI (eg. MRL, RPA, MMPR); if this would not give results fulfilling the requirements for the CCβ, a concentration between 50 and 100 % of the LoI should be chosen.

When possible, at least 20 different fortified blanks shall be analysed in order to ensure a reliable basis for this determination. When 20 different matrices are not available, it is recommended to test 5 replicates based on the same spiked matrix (e.g. 20 replicates at least 4 different matrices; 40 replicates at least 8 different matrices; 60 replicates: at least 12 different matrices). This approach is already proposed in ISO technical specification 23758 [10].
The number of “Screen Positive” Control Samples (i. e. samples spiked at the STC) for each analyte depends on the degree of statistical confidence required in the result, and the relationship between the STC and the LoI. The lower the STC in comparison with the LoI, the fewer replicates are required to give the same degree of confidence that the screening test will correctly identify samples truly contaminated with concentrations at the LoI. It is recommended to analyse 20 to 60 samples depending on the relation between the STC and the LoI (**Table 6**).

Table 6. Recommended number of samples based on closeness of the STC to the level of interest (LoI).

Relationship STC/LoI	Number of samples required
STC ≤ 0.5 LoI	20
0.5 LoI<STC< 0.9 LoI	40
0.9 RL≤STC≤ LoI	60
> LoI	20

If the STC is set at half the LoI 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 it fulfills the requirements for CCβ as given in CIR (EU) 2021/808 [1] (β error ≤ 5%). As a consequence, CCβ is less than the LoI (eg. MRL) and even less than or equal to ½ LoI. If the sensitivity of the screening test is such that the STC approaches the LoI (10 % below the LoI), more "screen positive" control samples are required to demonstrate that CCβ is fit for the purpose. These larger studies can be undertaken sequentially, i. e. testing in groups of twenty control samples, and if more than one spiked sample falls below the cut-off level Fm, the validation can be abandoned at this point, the STC increased and the validation exercise repeated.
These validation experiments should be carried out on different days, preferably 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 the determination of CCβ is carried out under 'blind' conditions simultaneously with the determination of specificity (cf 5.4.1.1.) (i. e. the operators do not know which samples are spiked and which samples are blank). It is especially required when visual reading is applied. When a reader is used, blind coding is not necessary.

If the screening method is applicable to a single matrix of different animal species, the 20 to 60 different samples should also represent the different species (e. g. 20 porcine muscles, 20 bovine muscles, and 20 poultry muscles) (see section 5.4.2.).

An example of experimental plan with one validation standard at STC (Method 2) for qualitative and semi-quantitative methods without calibration curves (cases I and II (Table 1)) is proposed in **Annex 12.1.1.**
An example of experimental plan with several validation standard as potential STC (Method 2) for qualitative and semi-quantitative methods without calibration curves (cases I and II (Table 1)) is proposed in **Annex 12.1.2.**

5.3 Identification of Cut-Off level Fm and determination of CCβ

5.3.1 Method 2

This approach is a statistical approach which takes into account the β error of 5 % and is referred to in CIR (EU) 2021/808 [1] as "**Method 2**".
Validation of biological and biochemical screening methods (whether qualitative or semi-quantitative) or physicochemical methods (used as qualitative method) screening methods requires identification of a cut-off level at, or above which the sample is categorised as 'screen positive' and liable to physico-chemical confirmation (cases I and II as given in Table 1).

- In the case of a microbiological growth inhibition test, a typical cut-off level Fm would be an inhibition zone with a width of > x mm (e. g. 2 mm). In this case, any sample giving a zone of > 2 mm would be classified as 'screen positive'. All samples spiked at the STC should give zones > 2 mm to be classified as 'screen positive'.
- In the case of biochemical or physico-chemical methods, the approach for establishing cut-off levels Fm for semi-quantitative screening tests without a proper linear calibration is given in **Annex 12.1.3.**

A threshold level T and a cut-off level Fm can be calculated, both of which are matrix-specific. **Method 2** combined with the determination of a cut-off level Fm is most often applied to the validation of qualitative and semi-quantitative biological screening methods (i. e. Investigation of fortified blank

material at concentration levels at the STC) [11, 12]. However this method can also be applied to physico-chemical screening methods, which can be treated as qualitative [5], semi-quantitative or quantitative methods (see chapter 6.3.2.).

Threshold level T and cut-off level Fm have to be calculated from the analysis of at least 20 blank samples and at least 20 samples spiked at the STC, respectively.
An example of experimental plan for determination of specificity and detection capability according to **Method 2** is presented in Annex **12.1**..
Regarding the relationship between the concentrations of the analytes to be investigated and the signal (output) they trigger, two cases can be differentiated:

- **Proportional case (1): The signal is proportional to the concentrations (e. g. HPLC-DAD, LC-MS/MS) or cut-off levels calculated from concentrations.**

For biochemical tests used as semi-quantitative methods, threshold level T and cut-off level Fm can be calculated from the calculated concentrations of at least 20 blank and 20 spiked samples.
The assay of the day will be declared valid only if T is lower than Fm. The results of unknown samples will be declared negative if their result is lower than Fm. On the contrary, the results of unknown samples will be declared positive if their result is higher than Fm.

- **Inversely proportional case (2): The signal is inversely proportional to the concentrations (e. g. ELISA).**

For biochemical tests used as qualitative methods, the assay of the day will be declared valid only if T is higher than Fm. The results of unknown samples will be declared negative if their result is higher than Fm. On the contrary, the results of unknown samples will be declared positive if their result is at or below Fm.
Then the samples with responses *larger than* (1) or *lower than* (2) the cut-off level Fm are identified as *positive results*. Conversely, the spiked samples with results *lower than* (1) or *higher than* (2) the cut-off level Fm are identified as *negative results*.

If more than one spiked sample out of 20 (i. e. > 5 %) is negative, the STC chosen for the spiking study is too low as this STC would not reliably yield “screen positive” results. In this case:

- If STC was equal to the LoI, the validation study has to be abandoned for this analyte until the method has been improved.
- If STC was half of the LoI, the spiking concentration should be increased (e. g. to three quarters of the MRL) and the spiking study repeated.

Determination of CCβ: If a maximum of one of the 20 samples spiked at STC (i. e. ≤ 5 %) gives a “screen negative” result, the STC is adequate. The STC then equals the detection capability CCβ of the method.

5.4 Other validation parameters

5.4.1 Specificity/selectivity

Specificity is determined by the analysis of an appropriate number (n ≥ 20) of representative blank samples (different batches including different animal species if within the scope of the method (the matrices should be evenly divided)). The experiments for the determination of the specificity and of the threshold level T shall be performed simultaneously with the analysis of the samples spiked at the STC for the determination of CCβ.
Note: Blind analyses of blank (specificity) and spiked samples (CCβ) shall be preferably performed over several days, with different batches of kits when relevant (e. g. ELISA).

5.4.1.1 False positive rate

The false positive rate is a percentage, calculated as the ratio between the number of negative results wrongly categorized as positive (false positive) and the total number of actual negative results.

Usually it is the interest of the laboratory to keep the false positive rate at a reasonable level in order to avoid unnecessary subsequent confirmatory analysis. Hence there is no specification with regard to an acceptable false positive rate.

5.4.1.2 Cross-reactions (CR) for biochemical tests

For a targeted screening test designed to selectively detect a single compound or an antimicrobial family (i. e. ELISA kit, receptor assay, etc.), the specificity/selectivity is also analysed by determining the rate of cross-reactions (CR) between the main analyte (MA) (detected at 100 %) (reference) and potentially interfering (PI) substances. For the conduction of the experiments on potentially interfering substances, chemically related substances (metabolites, derivatives, etc.) or other substances likely to be present in the samples (veterinary drug in combination with other drugs) are selected and spiked in appropriate blank samples, to analyse their influence on the test results. Blank samples should be spiked with these analytes at high concentrations (eg. 100 times the MRL (or MMPR, RPA)).
Spiked concentrations could be selected on the basis of supplier data, regulatory limits (if available) or in the absence of data by a preliminary assessment of potentially interfering analytes. When a supplier of a commercial kit (ie. ELISA) claimed in the leaflet or the validation report a low CR (eg. less than 1%) for a substance, the tested concentration should be high, for example 10000 µg/kg. Different concentrations could be tested for one potentially interfering substance.
To determine the rate of cross-reactions, at least 3 samples for each combination potentially interfering substance/concentration shall be tested the same day. Each of these blank samples must also be analysed on the same day. In parallel, the calibration curve, at least two screen negative controls (blank samples) and at least two screen positive controls (samples spiked with the analyte detected by the kit at CCβ) are analysed.
When high CR is calculated, the detection capability of the compound should be determined.
CR are determined in the concerned matrix. If the applicability of the method has been proved for other matrices, the CR have to be determined in one representative matrix.

Determination of cross-reactions
The calibration curve is used to calculate the corresponding concentrations of the tested samples. The QC are used to validate the assay of the day, and to calculate the recovery. The response of each potentially interfering substance is compared to a calibration curve of the main analyte to determine the concentration equivalent to the "main analyte".

Interpretation of the results of cross-reactions:
The analyte concentrations are usually calculated automatically by the software, by comparison with a calibration curve. Then, from the main analyte (detected at 100%), the average recovery of extraction is calculated as follows:

(1) $$\boldsymbol{Recovery} = \frac{(\bar{x}_{MA,spike} - \bar{x}_{MA,blank})}{Spiked\ concentration_{MA}} * \mathbf{100}$$

Where:
$\bar{x}_{MA,spike}$: mean calculated concentration of the main analyte (MA) in the spiked sample;
$\bar{x}_{MA,blank}$: mean concentration (equivalent main analyte) in the blank sample;
Spiked concentration $_{MA}$: concentration of the main analyte spiked in the sample.

For potentially interfering substances (PI), the difference between the mean concentration calculated for the samples spiked with PI substances and the mean concentration calculated for blank samples is calculated.

(2) **Difference =** $(\bar{x}_{PI,spike} - \bar{x}_{blank})$

Where:
$\bar{x}_{PI,spike}$: mean concentration of samples spiked with PI substances in equivalent MA;

$\bar{x}_{blank}$: mean concentration of blank samples in equivalent MA.

Finally, the recovery is applied to all antibiotics detected by the same test as follows:

(3) Calculated concentration of the sample spiked with PI substances:

Calculated [PI]=$\frac{(\bar{x}_{PI,spike}-\bar{x}_{blank})}{Recovery} * 100$

Then, the percentage of cross-reactions of all PI analytes (%CR_{PI}) is calculated as follows:

(4) $$\%\boldsymbol{CR_{PI}} = \frac{\boldsymbol{Calculated\ [PI]}}{\boldsymbol{Spiked\ [PI]}} * \mathbf{100}$$

Where:

Calculated [PI]: Calculated concentration of the sample spiked with PI substances;
Spiked [PI]: Concentration of PI substance spiked in the sample.

Estimation of CCβ for molecules that have been validated only by cross-reactions studies and not by determination of CCβ as described in paragraph 5.3.

There are two possible scenarios:

- If CCβ has been determined for the main analyte (CR 100%), the CCβ of the potentially interfering analyte can then be estimated as follows:

(5) $$\boldsymbol{CC\beta_{PI}} = \frac{\boldsymbol{CC\beta_{MA}}}{\%\boldsymbol{CR_{PI}}\%} * \mathbf{100}$$

Where:

$CC\beta_{PI}$: CCβ of the PI analyte;
$CC\beta_{MA}$: CCβ of the main analyte;
%CR_{PI}: Cross-reaction percentage of the PI analyte.

- If CCβ has been determined for an interfering substance (most often the least well detected of the desired family), it is first necessary to determine the CCβ of the main analyte (% CR = 100), for which in this case we have probably only determined the cross-reactions. In this case, we will apply the following formula:

(6) $$\boldsymbol{CC\beta_{MA}} = \frac{(\boldsymbol{CC\beta_{PI}} * \%\boldsymbol{CR_{PI}})}{\mathbf{100}}$$

Where:

$CC\beta_{MA}$: CCβ of the main analyte;
$CC\beta_{PI}$: CCβ of the PI analyte;
%CR_{PI}: Cross-reaction percentage of the PI analyte.

5.4.2 Method applicability

The applicability consists in demonstrating during the full validation that the method is applicable to different animal species or matrices with the same detection capabilities.
The applicability of a newly developed screening method to different matrices (and/or different animal species) should be demonstrated by the determination of specificity and detection capability CCβ for these different matrices.
At least 20 blank samples (5 samples per combination matrix/species) and the same 20 blank samples spiked at the STC shall be analysed.
A common specificity and common CCβ for the tested analytes would be determined if less than 5 % of the spiked sample are negative.
The detection capability CCβ could be overestimated for one animal species for instance because a common CCβ is determined (ie. A common STC). For instance, a common CCβ could have been

determined at ½ MRL (if same MRL for both animal species), while a lower CCβ could have been reached for one of the two animal species.

5.4.3 Stability

"If stability data for analytes in the matrix are available (e. g. on the basis of information from the EURLs, published data, etc.), these data do not need to be determined by each laboratory. However, referring to available stability data of analytes in solution and in matrix is only acceptable if identical conditions are applied." (CIR (EU) 2021/808 [1]).
For biological and biochemical methods, stability studies are generally not determined in the concerned laboratories. It is recommended to use stability studies implemented with physico-chemical methods available in the bibliography.

5.4.4 Ruggedness

Ruggedness studies focus on 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 annex of CIR (EU) 2021/808 [1], by means of experimental plans. "Each performance characteristic shall be determined for all minor changes that have been shown to have a significant effect on the performance of the assay" [1].
The ruggedness of the method should preferably be determined prior to validation.
Matrices (eg. Matrix composition, matrix quality) 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 (ie. specificity) and 10 different materials spiked (or incurred) at the level of interest (ie. STC) (detection capability CCβ). It is recommended to perform the ruggedness studies as a blind test (unknown samples). For parameters of the test protocol (eg. Incubation temperature or time), 4 blank and 4 spiked samples are sufficient.
When it has been demonstrated that one factor gives 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 method.

One approach to this issue is to use the **factorial approach** as described in Chapter 2.2.2. ("Validation according to alternative models) of the CIR (EU) 2021/808 [1].

6 Validation with calibration curves (Cases III and IV)

6.1 Principle

For semi-quantitative and quantitative methods with calibration curves, the performance characteristics to be determined are detection capability CCβ, specificity/selectivity, trueness, precision, relative matrix effect (if internal standards are used), and absolute recovery (if internal standards are not used).
Trueness, precision, relative matrix effect, and absolute recovery are described in chapter 6.4.2.

6.2 Samples and Experiments

For the validation experiments, a number of different blank and spiked samples at STC-level is required, which are prepared according to the method description of the analytical method to be validated.

The STC at which the matrix blank samples will be spiked should ideally be set at half the level of interest (e. g. MRL, RPA, MMPR); if this would not give results fulfilling the requirements for the CCβ, a concentration between 50 and 100 % of the level of interest should be chosen.
Regarding the minimum number of samples required for the full validation of a screening the recommendations given in the EURL Guidance on Confirmation Method Validation and/or the guidelines given in chapter 5.2 can be used for orientation. Methods 1 and 3 for the calculation of CCβ do not provide a minimum sample amount, whereas if method 2 is to be applied a minimum of 20 samples need to be analysed in order to prove that the 5 % β-error can be respected. As for (semi-)quantitative methods many of the validation parameters are the same as for confirmation methods, the model validation plans for the conventional or alternative validation concept given in the EURL Guidance on confirmatory method validation (in Tables 3 to 5). can be followed for the validation of screening methods In contrast to a confirmation method all obtained measurement results can be included in the evaluation regardless of whether or not the analyte of interest could be identified unequivocally. For LC-MS/MS methods it would therefore usually be sufficient to include only the main transition in the validation.

6.3 Determination of CCβ

The CCβ can be determined based on (semi-)quantitative analytical results which is usually done for physico-chemical methods and described in **Methods 1, 2 and 3** given in CIR (EU) 2021/808 [1].

6.3.1 Method 1

The calibration curve procedure according to ISO 11843-1:1997 (here referred to as minimum detectable value of the net state variable).

- Unauthorised or prohibited pharmacologically active substances: In this case, representative blank material shall be used, which is fortified at and below the RPA, or if no RPA has been established, around the STC in equidistant steps.
- Authorised substances: In this case, representative blank material shall be used, which is fortified at and below the permitted limit, starting from the STC in equidistant steps.

The corresponding concentration at the STC plus 1.64 times the standard deviation of the within-laboratory reproducibility of the mean measured content at the STC equals the detection capability.

For practical implementation, an example of experimental plan is proposed in **Annex 12.2.1.**

6.3.2 Method 2

The CCβ can be calculated according to **method 2** for semi-quantitative and quantitative physico-chemical screening methods with calibration curves, as well as for biochemical methods (eg. ELISA) (Cases III and IV of **Table 1**):

For practical implementation, an example of experimental plan is proposed in **Annex 12.1.2.**

CCβ is assessed from the fortified samples at the different validation levels. For each concentration level, the 20 fortified blanks are analysed.

The detection capability of the method is equal to the concentration level where only 5% or less false conforming results remain [1].

Thus :

- During validation, the STC is determined with an S/N close to 10 on the most abundant signal;

- At the CCβ level, no analytical performance is required other than that described in Method 2 for the determination of CCβ (suspected 19 times out of 20, with the chosen screening identification criteria).

It is up to the laboratory ("laboratory policy") to choose an appropriate threshold from which samples are sent to confirmation. For authorised substances, it could be ½ MRL (eg. For control plans) or 0.1 MRL (eg. For surveillance plans). For banned substances, it could be the STC or even each sample for which the signal to noise ratio is higher than three (S/N>3) for 1 or 2 transitions.

6.3.3 Method 3

The principle of the calculation is shown in the following graphics (**Figure 2**). As a result of the validation a calibration curve and a respective confidence interval is calculated. Using this function, a maximum STC (STC_{max}) can be calculated which still fulfils the requirement for a screening method (CCb <= MRL/RPA/MMPR).

$$\mathbf{STC_{MAX} = MRL - k * u_{(MRL)}}$$

In general, the STC refers to a true/spiked concentration and corresponds to a measured value. The STC_{max} corresponds to a maximum Decision value (DV_{max}). In addition a minimum decision value (and a corresponding minimum STC) can be calculated based on the results of the analysis of blank samples. It is now up to the laboratory ("laboratory policy") to chose an appropriate DV (between DV_{min} and DV_{max}) from which on samples are sent to confirmation. With this the fulfilment of the requirement for a screening method is guaranteed.

The practical approach for determination of specificity and detection capability according to Methods 1 and 3 is presented in **Annex 12.2.2.**

CCβ = STC + k(one-sided, 95 %) × (combined) standard measurement uncertainty at or above the STC.

For practical implementation, an example of experimental plan is proposed in **Annex 12.2.1.**

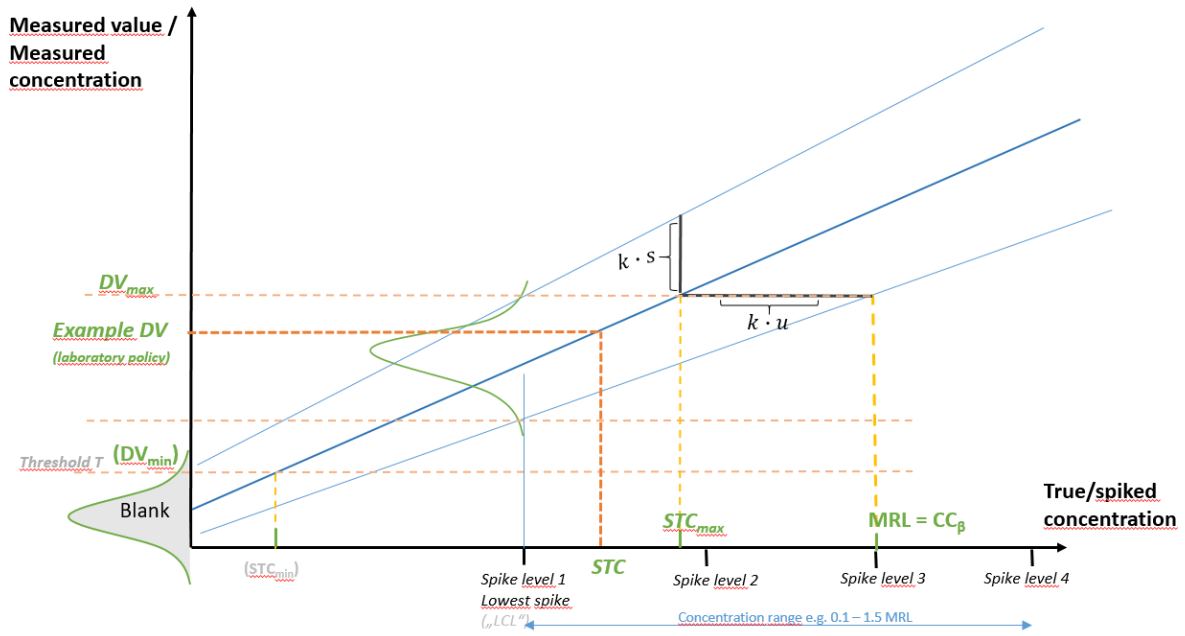


Figure 2. Link between CC_{β}, Screening target concentration (STC) and Decision value DV (STC_{max} is chosen in order to get CC_{β} = MRL).

6.4 Other validation parameters

6.4.1 Specificity/selectivity

The specificity/selectivity is the power of discrimination between the analyte and closely related substances. It should be determined as described in chapter 2.3. of CIR (EU) 2021/808 [1].

6.4.2 Trueness, precision, relative matrix effect, absolute recovery

Trueness and precision should be determined as described in chapter 2.2. of CIR (EU) 2021/808 [1].
Relative matrix effect and absolute recovery are relevant for quantitative screening methods to prove by means of the validation that the requirements for the performance characteristics are met. The relative matrix effect of the method shall be determined when this effect was not assessed during the validation procedure. The absolute recovery of the method shall be determined when no internal standard or no matrix-fortified calibration is used.
The precision requirements of Chapter 1.2.2.2 of CIR (EU) 2021/808 do not need to be met for semi-quantitative screening methods. However, the precision shall be determined to prove the suitability of the method for avoiding false compliant analytical results.

The precision requirements of Chapter 1.2.2.2 of CIR (EU) 2021/808 need to be met for quantitative screening methods.

6.4.3 Stability

See 2.5, Annex 1 of Commission Implementing Regulation (EU) 2021/808, experiments for stability are not included in the exemplary validation plans given in this guidance.

6.4.4 Ruggedness

No fixed criteria. The results for the parameter shall be evaluated using expert knowledge. The responsible scientist shall identify critical aspects which may require method improvements.

7 Fitness for purpose

When the method is only used for screening purposes, the requirements for identification (cf. 1.2.3, 1.2.4, 1.2.5) of the CIR (EU) 2021/808/EC [1] do not need to be fulfilled.

Regardless of which validation approach has been used it is necessary to assess the outcome of the validation study for the method's fitness for purpose and address all relevant aspects in the validation report. The acceptance criteria by which to judge whether or not a screening method can be considered adequately valid for a certain substance are given in **Table 4**.
If any of these criteria are not met for a substance, the method is not fit for the intended purpose. The consequences would be

- to define the applicability of the method accordingly (e. g. only applicable for the detection for 3 out of 4 initially intended substances)
- to define the method for a different purpose (e. g. only qualitative screening if quantitative requirements are not fulfilled) if the required data is available. In cases where the validation data implies that an analytical method does not fulfil all requirements for quantitative screening methods as laid down in CIR (EU) 2021/808 [1], the method may still by applicable as a qualitative screening method.
- to continue method development followed by another attempt at method validation.

8 Validation report

When a screening method has been validated, a report of the validation study has to be compiled.

As described in the standard ISO 17025 [13], the performance characteristics of the validated methods must meet the specified requirements. In addition, the laboratory must keep the following validation records: the validation procedure used, specification of requirements, determination of the performance characteristics of the method, the results obtained and a statement of the validity of the method (cf. 7.), giving details of its performance.

The full validation report should contain at least:

- identify the application range of the method, including the ruggedness statements, concentration range (for (semi)-quantitative methods), 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.

9 Ongoing method performance verification

As it is described in the “EURL Guidance Document on the Quality control during routine analysis (ongoing method performance verification)”, the analysis of QC samples and the continued evaluation of calibration curves are valuable tools for the provision of evidence of method performance in routine analysis. This guidance on the quality control during routine analysis may be used – partly with adaptation – for screening methods.
This chapter supplies additional information specific for screening methods.

Regardless of detection principle, the analysis of QC samples is of the utmost importance as it enables the laboratories to identify problems during analyses and assists in safeguarding a method’s correct application. The use of spiked samples as QC is applicable to qualitative tests (e. g. tube tests, receptor-based tests), semi-quantitative (e. g. ELISA kits) and quantitative methods (e. g. LC-MS/MS methods). For microbiological plate tests, it is highly preferable to use incurred samples where possible, or spiked 'simulated tissue' slices as used during the validation phases. If such materials are not available, at least paper discs spiked with antibiotic standard (for each plate) shall be used as positive QC samples.

For biochemical tests (eg. ELISA kits), each batch of analyses should include both screen negative control sample (at least two matrix blank samples) and screen positive control samples (at least two matrix blank samples fortified at the CCβ of validation). The threshold level T and the cut-off level Fm should be calculated every day from the blank samples and from the spiked samples at CCβ respectively. If the screen positive control sample gives a negative result (i. e. less than the cut-pff level Fm), the batch of analyses should be discarded. Similarly, if the screen negative control samples give a positive result (i. e. above the cut-off level Fm), the batch of analyses should be discarded. In both cases, there should be an investigation into why the test has failed and remedial action should be taken. Results from the QC samples should be recorded continuously (e. g. in control charts) 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.

For ELISA tests used as qualitative tests, QC samples shall be used to determine the daily threshold level T (from matrix blank (screen negative control sample)) and cut-off level Fm (from fortified matrix blank ("sreen positive control sample)). Each day of analysis, T and Fm values have to be calculated from the analysis of at least 2 screen negative control samples and 2 screen positive control samples. The assay of the day will be declared valid only if T is higher than Fm. The results of unknown samples will be declared negative if their result is higher than Fm. On the contrary, the results of unknown samples will be declared positive if their result is lower than Fm.

Note: The choice of analytes to be included in routine QC samples should follow the same rules as those selected for the initial 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.

10 Extension of methods

The experiments 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.

As it is described in the "EURL Guidance document on the extension of quantitative confirmation methods", sometimes it becomes necessary to extend the scope of a previously comprehensively validated method (confirmatory or screening method). An extension of the scope should be accomplished in an efficient and analytically sound way, ideally using a reduced number of samples compared to a full validation. This guidance document on the extension of quantitative confirmation methods may be used – partly with adaptation - for screening methods.

For qualitative screening methods (or (semi-)quantitative methods validated in terms of a qualitative screening method), when the applicability has to be adapted after the initial validation, the experiments correspond to an extension of method. The extension can apply to new matrices, species, substances, and concentrations.

- New matrix/species: At least 20 blank samples for the new combination matrix/species and the same 20 blank samples spiked at the STC (same STC used for the original matrix during the initial validation) shall be analysed (**Figure 3**). A common specificity and common CCβ (i. e. parameters from initial validation) for the tested analytes would be determined if less than 5 % of the screen positive control samples yield negative results. If more than 5 % of the screen positive control samples give negative results, the method is not applicable with the same CCβ. A full validation has to be done for the new matrix.

- New analytes: At least 20 blank samples and the same 20 blank samples spiked at the STC with the new analyte shall be analysed. The STC is adequate and the CCβ for the tested analyte can be determined (CCβ = STC) if less than 5 % of the screen positive control samples give negative results. If more than 5 % of the screen positive control samples give negative results, the STC has to be increased or a full validation has to be done for the new analyte.

- New concentrations: If the level of interest is changed, it could be necessary to determine a new detection capability CCβ if the previous CCβ is now higher than the new LoI. For instance if the MRL is decreased and the new MRL differ greatly, it might be necessary to determine a new CCβ. It is necessary when method 2 was used to determine detection capability. In this case, at least 20 blank samples and the same 20 blank samples spiked at the new concentration shall be analysed. The CCβ for the tested analytes would be determined if less than 5 % of the screen positive control sample give negative results. If more than 5% of the screen positive control sample give negative results, the concentration has to be increased. The CCβ will be valid if it is lower than or equal to the new LoI.

If the LoI is increased or if the modification of the LoI is very low, it might not be necessary to determine a new CCβ.

Note: the monitoring of the method performance through ongoing quality control is required to collect the data missing for a complete validation.

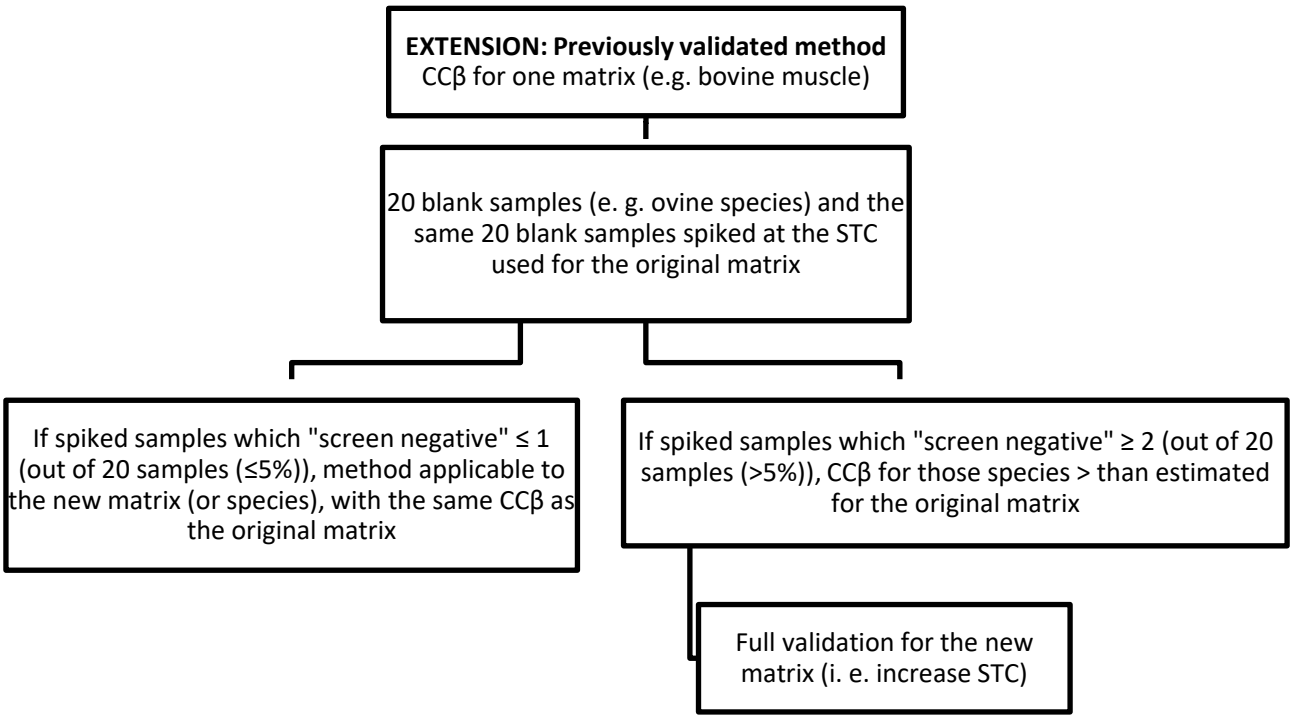


Figure 3. Extension of qualitative screening methods and screening method applicability to other matrices/animal species: Decision tree.

11 References

[1] *Commission Implementing Regulation (EU) 2021/808 of 22 March 2021 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals and on the interpretation of results as well as on the methods to be used for sampling and repealing Decisions 2002/657/EC and 98/179/EC (Text with EEA relevance).* Official Journal of the European Union, 2021. **L 180**: p. 84–109

[2] *Commission Implementing Regulation (EU) 2018/470 of 21 March 2018 on detailed rules on the maximum residue limit to be considered for control purposes for foodstuffs derived from animals which have been treated in the EU under Article 11 of Directive 2001/82/EC (Text with EEA relevance.).* 2018: Official Journal. p. p. 16–18

[3] *EURL GUIDANCE ON MINIMUM METHOD PERFORMANCE REQUIREMENTS (MMPRs) FOR SPECIFIC PHARMACOLOGICALLY ACTIVE SUBSTANCES IN SPECIFIC ANIMAL MATRICES.* 2022. **June 2022**.

[4] *Commission Regulation (EC) N° 470/2009 laying down Community procedures for the establishment of residue limits of pharmacologically active substances in foodstuffs of animal origin* 2009, The European parliement and the Council: Official Journal of the European Union p. 11-22.

[5] Gaugain-Juhel, M., B. Delépine, S. Gautier, M.P. Fourmond, V. Gaudin, D. Hurtaud-Pessel, E. Verdon, and P. Sanders, *Validation of a liquid chromatography-tandem mass spectrometry screening method to monitor 58 antibiotics in milk: a qualitative approach.* Food Additives & Contaminants: Part A, 2009. **26**(11): p. 1459-1471.

[6] Gaudin, V., C. Hedou, A. Rault, P. Sanders, and E. Verdon, *Comparative study of three screening tests, two microbiological tube tests, and a multi-sulphonamide ELISA kit for the detection of antimicrobial and sulphonamide residues in eggs.* Food Additives & Contaminants: Part A, 2009. **26**(4): p. 427–440.

[7] Gaudin, V., P. Maris, R. Fuselier, J.L. Ribouchon, N. Cadieu, and A. Rault, *Validation of a microbiological method: the STAR protocol, a five-plate test, for the screening of antibiotic residues in milk.* Food Additives and Contaminants, 2004. **21**(5): p. 422-433.

[8] Gaudin, V., C. Hedou, A. Rault, and E. Verdon, *Validation of a Five Plate Test, the STAR protocol, for the screening of antibiotic residues in muscle from different animal species according to European Decision 2002/657/EC.* Food Additives & Contaminants: Part A, 2010. **27**(7): p. 935–952.

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

[10] 2021, *ISO/TS 23758 Technical Specification "Guidelines for the validation of qualitative screening methods for the detection of residues of veterinary drugs in milk and milk products"*

"Lignes directrices pour la validation des méthodes qualitatives de dépistage des résidus de médicaments vétérinaires dans le lait et les produits laitiers".

[11] Gaudin, V., C. Hédou, A. Rault, E. Verdon, and C. Soumet, *Evaluation of ELISA kits for the screening of four nitrofuran metabolites in aquaculture products according to*

the European guideline for the validation of screening methods. Food Additives & Contaminants: Part A, 2021. **38**(2): p. 237-254.

[12] Gaudin, V., C. Hedou, C. Soumet, and E. Verdon, *Multiplex immunoassay based on biochip technology for the screening of antibiotic residues in milk: validation according to the European guideline.* Food Additives & Contaminants: Part A, 2018. **35**(12): p. 2348-2365.

[13] 2005, *ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories*.

12 Annexes

12.1 Validation of screening methods without calibration curves (cases I and II (Table 1))

In the following two examples (12.1.1 and 12.1.2) the detection capability CCβ refers to the STC.

12.1.1 Validation of qualitative and (semi-)quantitative methods with one validation standard at STC (Method 2)

A minimum of 20 different batches of matrices are needed in a full validation for the determination of specificity and detection capability CCβ. These 20 samples can be spread over 4 days for example (Table 7).
The calibration curve is not necessary for this experimental plan. For instance, microbiological methods do not include a calibration curve. Biochemical methods (eg. ELISA, biosensors) could be validated following this general scheme with or without analysing a calibration curve.

Table 7. Experimental plan for determination of specificity and CCβ with Method 2.

Validation series 1	5 blank samples for specificity and 5 spiked samples at the STC + calibration curve (if relevant)
Validation series 2	5 blank samples for specificity and 5 spiked samples at the STC + calibration curve (if relevant)
Validation series 3	5 blank samples for specificity and 5 spiked samples at the STC + calibration curve (if relevant)
Validation series 4	5 blank samples for specificity and 5 spiked samples at the STC + calibration curve (if relevant)

All of these batches, except the batch used for the calibration curve, need to be fortified with the analytes of interest at the STC. Regarding commercial kits (eg. ELISA), the calibration standards are most often prepared in buffer, not in the matrix.
The choice of the STC depends on the legal status of the residue in question and on the estimated sensitivity of the screening method (eg. Detection limit announced by the manufacturer of a commercial kit).
If less validation series are performed (Table 8), it is preferable to space out series (e.g. one round per week).

Table 8. Experimental plan in relation to the number of samples required.

Relationship STC/LoI	Number of samples required	Practical approach*
STC ≤ 0.5 LoI	20	Day 1 5 samples, day 2 5 samples, day 3 5 samples, day 4 5 samples - blank and spiked
0.5 RL<STC< 0.9 LoI	40	Day 1 10 samples, day 2 10 samples, day 3 10 samples, day 4 10 samples - blank and spiked
0.9 RL≤STC≤ LoI	60	Day 1 15 samples, day 2 15 samples, day 3 15 samples, day 4 15 samples - blank and spiked
> LoI	20	Day 1 5 samples, day 2 5 samples, day 3 5 samples, day 4 5 samples - blank and spiked

**if this is too much for one day, then divide over 3 days*

In this case, the samples can simply be evaluated as "screening positive" and "screening negative" (regardless of whether quantification would also be possible).
According to 6.3.2 the detection capability of the method is equal to the concentration level where only 5% or less false conforming results remain. I.e. if the required number of samples given in table 8 fulfills this criterion, the STC = CCβ.

12.1.2 Validation of qualitative and (semi-)quantitative methods with several validation standards as potential STC (Method 2)

Method 2 for the determination of detection capability CCβ is also applicable to semi-quantitative and quantitative methods with calibration curves. One of the calibrants is generally the LCL (lowest calibration level) as the relevant STC for a screening purpose.
CCβ is assessed from the fortified samples at the different validation levels (**Table 9**).

Table 9. Required fortification levels in a conventional validation study as given in CIR (EU) 2021/808 [1].

Residue	Level 1	Level 2	Level 3
Unauthorised with RPA[4]	0.5[5]·RPA	1.0·RPA	1.5·RPA
Unauthorised with MMPR	0.5[6]·MMPR	1.0 MMPR	1.5·MMPR
Unauthorised	1.0·LCL	2.0 LCL	3.0·LCL
Authorised (with MRL)	0.1[7]·MRL/ML	1.0·MRL/ML	1.5·MRL/ML
Authorised (without MRL in the matrix of interest) [8]	≤0.125·cascade MRL[9]	0.25 cascade MRL	cascade MRL

It is possible to extend the concentration ranges (e.g. level 1 < 0.1 MRL or level 3 > 1.5 MRL) or to add an additional intermediate level that seems important to characterise the analytical method (e.g. 0.5 MRL) and ensures a sufficient validation range in case of non-compliance with the validation criteria at level 1.

A general validation scheme is presented in **Table 10**.

[4] The concentration levels given here for RPA substances are to be understood as exemplary concentration levels. Analytical methods for the confirmation of substances for which an RPA has been established shall be validated at concentrations as low as reasonably achievable.
[5] Where 0.5 RPA is not reasonably achievable, this level can be replaced by the lowest reasonably achievable concentration between 0.5 and 1.0 RPA.
[6] Where 0.5 MMPR is not reasonably achievable, this level can be replaced by the lowest reasonably achievable concentration between 0.5 and 1.0 MMPR.
[7] Where 0.1 MRL/ML is not reasonably achievable, this level can be replaced by the lowest reasonably achievable concentration between 0.1 and 0.5 MRL/ML.
[8] For authorised pharmacologically active substances, for which there is no MRL in a specific matrix or species, the MMPR is equal to 0.25 MRL (cascade), established under Regulation (EU) 2018/470 for the substance concerned, where analytically feasible.
[9] If the sensitivity of the analytical method allows and for practical reasons of harmonisation with other substances, it will be appropriate to validate at the 0.1 cascade MRL level, the aim being to ensure a CCα < MMPR.

Series 1 allows to determine specificity and matrix effect (intra-matrix/specie and inter-matrix/specie) covering the application field of the method. Coefficient of variations (CV) are calculated to evaluate matrix effects.

- if CV is lower than 20%, the matrix/specie effect is not significant; Series 2, 3 and 4 can be homogeneous;
- if CV is higher than 20%, the matrix/specie effect is significant; Series 2, 3 and 4 are targeted onto one type of matrix/specie. For each run, the different batches will be from the same species/matrix.

The batch used for the calibration standards should be different for each run.
If equipment availability allows, it is preferable to space out series 2, 3 and 4 (e.g. one round per week). Re-injection of one of the runs after 24, 48 or 72 hours allows the stability of the analytes in the extracts to be checked before injection. Only the range and validation standards corresponding to the level of interest can then be re-injected.

Table 10. Example of general validation scheme with different validation levels (Method 2).

Validation series 1	/	20 different batches of matrix: blank and spiked (Level 2)
Validation series 2	Calibration curve (n≥5)	n≥6* (different batches) Blanks** Level 1 Level 2 Level 3
Validation series 3	Calibration curve (n≥5)	n≥6* (different batches) Blanks** Level 1 Level 2 Level 3
Validation series 4	Calibration curve (n≥5)	n≥6* (different batches) Blanks** Level 1 Level 2 Level 3

**If specificity was not evaluated during series 1, 7 replicates are needed, instead of 6.*
***It is preferable to analyse the different batches of blank samples over the 3 different series.*

Determination of CCβ. If a maximum of one of the 20 samples spiked at STC (i. e. ≤ 5 %) gives a "screen negative" result, the STC is adequate. The STC then equals the detection capability CCβ of the method.

12.1.3 Validation of (semi-)quantitative methods by using the response information in a semi-quantitative way for the calculation of threshold level T, cut-off level Fm (Method 2)

This approach can be used for analytical methods with provides a somehow quantitative response and allows a more statistically sound calculation of the method performance compared to 12.1.1.
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.
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 level" Fm can be calculated.

Threshold T and cut-off level Fm are matrix-specific.

Two cases shall be presented:

- **Case 1: The signal is proportional to the concentrations (e. g. HPLC, LC-MS/MS).**

Calculation of threshold value T:

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

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

Calculation of Cut-off level Fm:

$Fm = M - 1.64 \times SD$

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

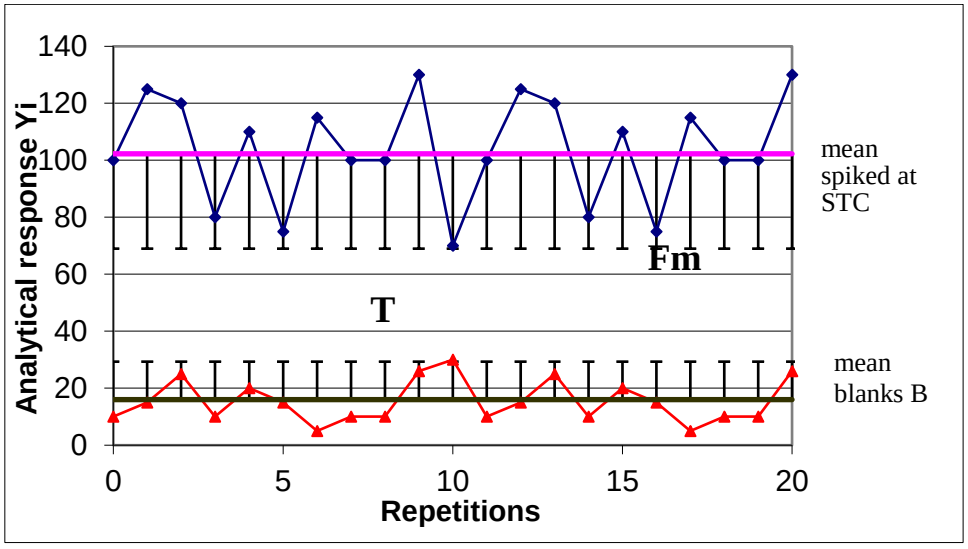


Figure 4. Graphical representation of threshold level T and "Cut-Off" Level Fm.

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

According to the CIR (EU) 2021/808 [1], the chosen STC and hence the detection capability CCβ are valid if : Fm > T. In this case, the false negative rate is lower than or equal to 5% and CCβ = STC. **Determination of CCβ. If a maximum of one of the 20 samples spiked at STC (i. e. ≤ 5 %) gives a "screen negative" result, the STC is adequate. The STC then equals the detection capability CCβ of the method.**

Also the laboratory should determine the rate of false positive (FP) and specify what is acceptable with the method. There is no specification with regard to an acceptable false positive rate. When the analytic response (or concentration) is higher than T, the rate of FP is below 5 %.

- **Case 2: The signal is inversely proportional to the concentrations (e. g. ELISA).**

For ELISA tests, the response (B/B0 %) is inversely proportional to the concentration, Therefore: Calculation of threshold value T based on the response:

T = B - 1.64xSDb or technical threshold.

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

Calculation of Cut-off factor Fm:

$$Fm = M + 1.64 \times SD$$

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

According to the CIR (EU) 2021/808 [1], the chosen STC and hence the detection capability CCβ are valid if : Fm < T. In this case, the false negative rate is lower than or equal to 5% and CCβ = STC.
Determination of CCβ. If a maximum of one of the 20 samples spiked at STC (i. e. ≤ 5 %) gives a "screen negative" result, the STC is adequate. The STC then equals the detection capability CCβ of the method.

Also the laboratory has to determine the rate of false positive (FP) which is acceptable with the method. There is no specification with regard to an acceptable false positive rate. When the analytic response is lower than T, the rate of FP is below 5 %.

However, with ELISA tests, it is also possible to calculate T and Fm values on the calculated concentrations, instead of on the response (signal). In this case, the threshold values T and Fm are calculated as for physico-chemical methods (case 1).

12.2 Validation of screening methods with calibration curves (Cases III and IV (Table 1))

In the following two examples (12.2.1 and 12.2.2) the detection capability CCβ can be calculated for a given STC based on the calibration function (cf. e.g. 6.3.). CCβ is assessed from the fortified samples at the different validation levels (**Table 9**).

12.2.1 Example of experimental plan for semi-quantitative and quantitative methods with calibration curves (cases III and IV) in accordance with the classical validation approach implementing Method 1 or 3

Methods 1 and 3 can be applied to semi-quantitative and quantitative methods with calibration curves for the determination of detection capability CCβ (see Table 1).
The following table proposes an experimental plan for the determination of CCβ when implementing method1 or method 3.

Table 11. Experimental plan for determination of specificity and CCβ with Methods 1 or 3.

Validation series 1	7 batches for specificity and fortification + 1 batch for the calibration curve
Validation series 2	7 batches for specificity and fortification + 1 batch for the calibration curve
Validation series 3	7 batches for specificity and fortification + 1 batch for the calibration curve and ruggedness

All of these batches, except the batch used for the calibration curve, need to be fortified with the analytes of interest to a minimum of the fortification levels given in **Table 9**; the exact concentrations depend on the legal status of the residue in question. Samples will also need to be fortified with internal standard solution, if applicable.
In case absolute recovery and matrix effect have to be determined, a fourth validation series may be necessary. Information on the matrix effect can partly also be gathered from preliminary experiments.

When method 1 is applied for the determination of detection capability:
The corresponding concentration at the STC plus 1.64 times the standard deviation of the within-laboratory reproducibility of the mean measured content at the STC equals the detection capability.

When method 3 is applied for the determination of detection capability:
CCβ = STC + k(one-sided, 95 %) × (combined) standard measurement uncertainty at or above the STC.

12.2.2 Example of a validation of quantitative and semi-quantitative methods (cases III and IV) implementing Method 1 or 3 in accordance with the factorial validation approach

The basic principles of validation according to the alternative approach are explained in detail in the "EURL Guidance Document on Confirmation Method Validation" and are not repeated here. It should be noted that for screening methods validated in accordance with the alternative validation concept all obtained results, regardless of whether the analytes could be identified unequivocally, should be included in the calculation of the method parameters.

In the following only an example of a study and a calculation of the performance characteristics is given.

Example (coccidiostats in egg):

7 factors were selected which might have an influence on the analytical result and which might be changed after the validation in the course of the application of the method. These factors can be varied on two levels each.

Table 7: Factors and factor levels

Factor	Factor level
Breeding	organic – conventional
Operator	routine – occasional
Status of matrix	fresh - lyophilised
Extract storage	injection into HPLC immediately – storage of injection solution for 2 – 3 days at -20 °C
SPE cartridges	Lot 1 – Lot 2
HPLC column	column 1 – column 2
Interruption of sample prep.	before the SPE – after the SPE

Factor 1 (Breeding) takes the different chicken husbandry systems into account: cage, floor and free-range husbandry as well as organic eggs (organic breeding and feeding conditions).

Factor 2 (Operator) was chosen to check the robustness of the method with regard to two different operators.

Factor 3 (Status of matrix, lyophilisation) was chosen because partially, the samples were stored or shipped lyophilised. In that case, instead of 2 g of fresh matrix, the corresponding amount of lyophilised material was used. By adding the correct amount of H_2O, the sample is reconstituted and 2 g of fresh material are obtained.

Factor 4 (Extract storage) was supposed to simulate the storage of the sample extracts before analysis. Storing the samples before the analysis may be necessary due to the temporary unavailability of the instrument (maintenance, still analysing other samples,…)

Factor 5 (SPE cartridges) was chosen to check the influence of different lots of solid-phase extraction materials.

Factor 6 (HPLC column) was chosen to check the robustness of the method when an old and a new column with a different serial number but the same material are used.

Factor 7 (Interruption of sample preparation) was supposed to simulate the storage of the sample extracts before the clean-up with the SPE cartridges. Storing the extracts before the clean-up may be necessary when the preparation steps are not finished yet because of the number of the samples to be prepared.

For the validation experiment, eight different sample preparation procedures (Run 01 – 08) has to be applied. The eight different factor level combinations follow the experimental design as described in CIR (EU) 2021/808. Each of the eight different runs consisted of different matrix samples. The different runs each included a blank matrix, blank matrix spiked with internal standards, 6 spiked samples, 6 calibration levels).

Table 8: Factor-level combinations for validation – Experimental design

Run	Factor-level combinations						
	Breeding	**Extract storage**	**Operator**	**Matrix**	**HPLC column**	**SPE cartridges**	**Interruption of sample preparation**
Run 04	Organic	Without	Routine	Fresh	Column 2	Lot 1	Before SPE
Run 06	Conventional	With (2-3 days, -20 °C)	Routine	Fresh	Column 1	Lot 2	Before SPE
Run 03	Organic	Without	Occasional	Fresh	Column 1	Lot 2	After SPE
Run 08	Conventional	Without	Routine	Lyo	Column 1	Lot 1	After SPE
Run 07	Conventional	Without	Occasional	Lyo	Column 2	Lot 2	Before SPE
Run 05	Conventional	With (2-3 days, -20 °C)	Occasional	Fresh	Column 2	Lot 1	After SPE
Run 02	Organic	With (2-3 days, -20 °C)	Routine	Lyo	Column 2	Lot 2	After SPE
Run 01	Organic	With (2-3 days, -20 °C)	Occasional	Lyo	Column 1	Lot 1	Before SPE

For validation purposes, specifically prepared mix solutions containing each of the analytes at the relevant concentration levels have to be prepared. The mixtures contain the analytes in the relevant concentration ranges, e. g. for substances with MRL or ML and substances which are not authorised as veterinary drugs. Examples here are clopidol (non authorised) and robenidine (ML of 25 µg/g).

Table 9: Concentration levels for validation (µg/kg).

Sample	Clopidol	Robenidine	e.g. substance XY with MRL 100
CL00	0.0	0.0	0.0
CL01	0.5	2.5	10.0
CL02	1.0	7.5	25.0
CL03	2.0	12.5	50.0
CL04	3.0	25.0	75.0
CL05	4.0	37.5	100.0
CL06	8.0	50.0	200.0

According to the factor level combinations for each of the eight runs blank matrix spiked at the concentration level in Table X is produced and analysed. The quantification is done either via a standard

calibration or via an independent matrix calibration curve. Subsequently for the eight runs a mean calibration curve and a confidence interval is calculated.

Depending on the experimental design the analysis of additional matrix blank samples may be required for the determination of specificity, selectivity.

In this example the total number of required samples would be:

20 blank samples (different egg samples)

6 * 8 = 48 samples (for the 8 runs)

6 * 8 = 48 samples (for the 8 calibration curves – either standard or matrix calibration).

The following figures give examples for an analyte with a regulatory limit and a non-authorised analyte.

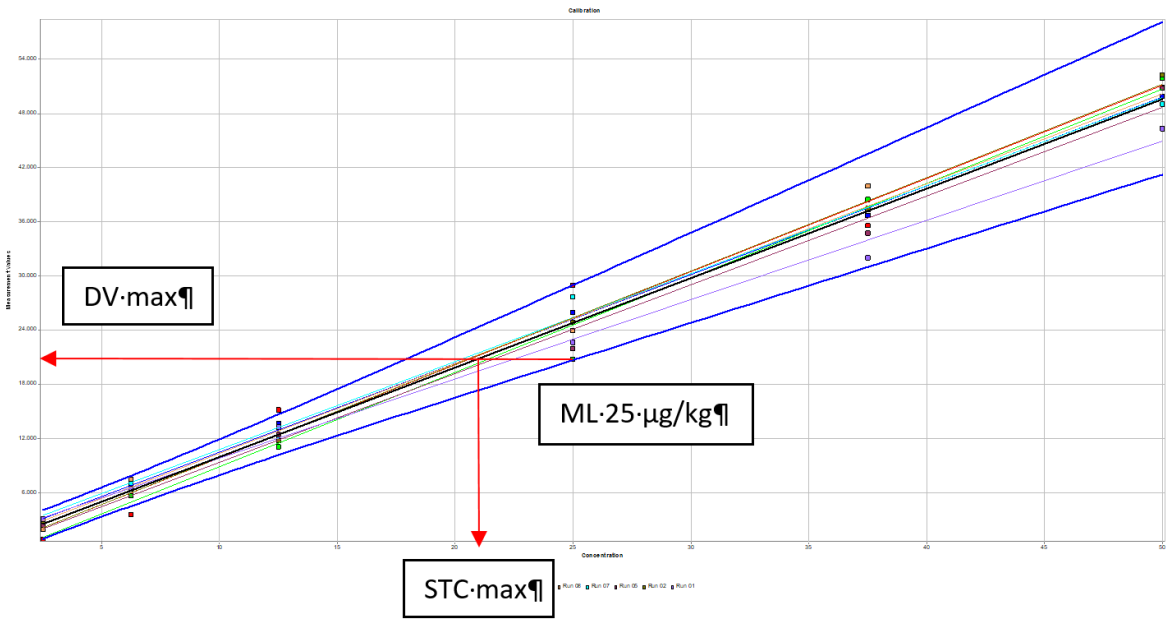


Fig. 5 Calibration curves for each run (results of the quantification of the fortification levels CL), overall calibration curve (black) and prediction interval (dark blue) for robenidine (ML 25 µg/kg). STC was selected as STCmax (highest STC so that the requirement CCβ <=ML is fulfilled).

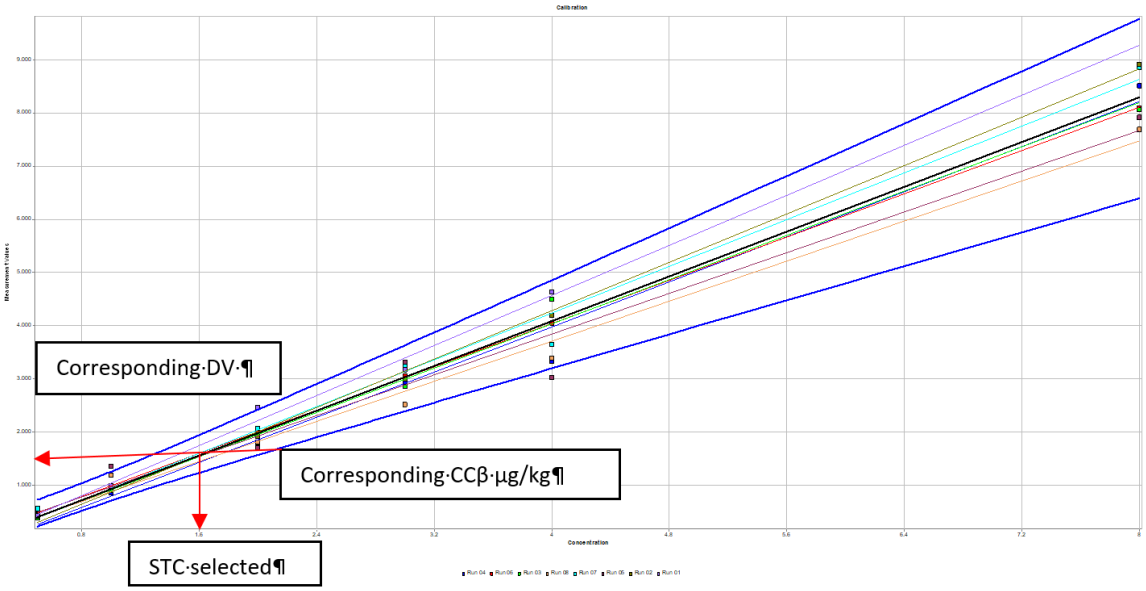


Fig. 6 Calibration curves for each run (results of the quantification of the fortification levels CL), overall calibration curve (black) and prediction interval (dark blue) for clopidol (not authorised). The STC was selected in the low calibration range ("conservative"), and the corresponding CCβ and DV were calculated accordingly.

CCβ = STC + k(one-sided, 95 %) × (combined) standard measurement uncertainty at or above the STC.

Alternatively the CCβ can be calculated using the calibration curve procedure (Method 1).