European Union
Reference Laboratories
WFSR and BVL Berlin
supported by the

Common EURL Protocol for Proficiency Testing in the Field of Veterinary Drug Residues

Table of contents

1. Introduction

This document serves as a general guideline for the conduction of proficiency tests (PT) in the function of European Union Reference Laboratory (EURL) by Wageningen Food Safety Research (WFSR) and Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) in the field of veterinary drug residues, henceforth referred to as "EURL". In addition to this guideline, the EURLs adhere to the principles laid down in ISO 17043, ISO 13528, and related standards. Specific cases may require a modification of this guideline which is within the responsibility of the respective EURL.

2. Organisation of Proficiency Tests

2.1. Objectives of PTs

The main objective of EURL PTs is to ensure reliable food control in the EU and partner states by:

- Allowing laboratories to check their individual performance
- Promoting the analysis of compounds of high significance
- Assessing method suitability
- Setting current performance benchmarks

2.2. Participants

The EURL PTs are primarily targeted towards the NRLs, which are obliged to participate according to Article 101, 1(a), Regulation (EU) 2017/625. The decision on the invitation of other EU official control laboratories (OCL) or laboratories from certain third countries is up to the EURL responsible for the respective PT. The organising EURL decides if handling fees will be charged for a PT.

2.3. Confidentiality

The participants consent to the publication of their results in an anonymous form by the organiser of the proficiency test. In the final report on results participants will only be referred to by their lab codes. The results of the NRLs participating in a proficiency test may be provided in a non-anonymous form to the European Commission (DG SANTE) for internal use, as well as to the competent authorities of the Member States (according to Article 94, 2(c), Regulation (EU) 2017/625). Analogously, the results of OCLs may be provided in a non-anonymous form to the responsible competent authorities for internal use (according to Article 101, 1(c), Regulation (EU) 2017/625).

3. Test Items

3.1. Selection of Samples

It is the goal of the EURLs to provide proficiency testing samples which resemble real-life samples and assess the laboratories' abilities to comply with current requirements. Therefore, the EURLs take the following aspects into account when deciding on number, type and amount of analytes:

- Legal status and suspected illegal use of veterinary drugs
- Number of livestock and relevant products in the member states
- Current developments (food scandals) and results of the National Residue Control Plans (NRCP)
- Relevant concentrations (e. g. MRLs)
- Challenging analytes, concentrations and matrices
- Technological advances (e. g. new measurement techniques)
- PTs as a means of assessing the suitability of EURL measures (e. g. collaborative trials)
- Suggestions from NRLs (e. g. during the annual EURL workshops)
- Sample availability

All decisions concerning the selection and preparation of the PT samples are made by the EURL organising the respective PT.

3.2. Homogeneity and Stability

In general, homogeneity and stability are assessed according to the standard procedures laid down in ISO 13528 Annex B or published by Thompson *et al.* as “The international harmonized protocol for the proficiency testing of analytical chemistry laboratories”. The decision on which procedure to use for homogeneity and stability assessment and on the inclusion of the uncertainties arising from inhomogeneity and insufficient stability for the calculation of the participant scores is within the responsibility of the respective EURL.

4. Communication with Participants

4.1. Information for Participants

Prior to the start of the PT participants receive information on the type and amount of test material, the compounds to be analysed, details on the statistical data analysis if necessary, modifications of the point score system (see section 5.8), as well as a timeline.

4.2. Analytical Requirements

If deemed necessary by the responsible EURL, specific requirements for the analytical results to be submitted by the participants (e. g. only three significant digits, two-fold analysis) are included in the PT protocol.

4.3. Method Details

Participants are asked to submit details on the applied sample preparation procedure and the analytical methods as well as certain performance characteristics.

4.4. Result Submission

The means of result submission (e. g. form, software, online tool) is decided by the responsible EURL. There is no restriction on how many times a participant may submit their results before the communicated deadline for submission. The final submission will be used as the base for all statistical calculations and the proficiency assessment. Results submitted after the deadline will in general not be used for the calculation of statistical parameters and will usually not be included in the PT report. Nevertheless, corrected participant results may be used to calculate material properties for a subsequent use of the remaining PT samples as in-house reference material.

4.5. Report on Results

Participants receive a final report on results which includes a compilation of all submitted results as well as all scores. Additionally, the PT provider may choose to prepare a preliminary report which compiles all submitted results and tentative scores shortly after the deadline for result submission so participants may promptly react to unsatisfactory results by implementing appropriate corrective measures. A preliminary report also serves the purpose of identifying any errors which occurred during data transfer or handling by the PT provider. Participants are therefore asked to carefully check their listed results and inform the PT provider of any discrepancies.

Errors may be corrected at any time, also after the publishing of the final report on results. In this case a corrected version of the report will be prepared.

5. Evaluation of Proficiency

5.1. Laboratory Results

Laboratory results x are the measurement results determined by means of the indicated analytical methods and submitted by the participants. In general, the results are accepted even if the methods are partly not validated, or if the validation data were not provided. The results are used as reported by the laboratories and the PT provider usually does not carry out any corrections, e. g. recovery corrections.

5.2. Assigned value x_{pt}

Due to the concept of free choice of method in the field of veterinary drug analysis, the assigned value is usually defined as the consensus mean obtained using robust statistics. When applying robust statistical methods, it is not necessary to exclude outliers as their influence on the consensus mean is neglectable. Still, the PT provider reserves the right to exclude participant results if a gross error can be proven or is reported by the participant. Likewise, results obtained using unsuitable analytical methods may be removed for calculation of the PT parameters. Other methods for assigning x_{pt} may be used if they are in compliance with ISO 17043. The PT provider will include information on the mode for assigning x_{pt} in the final report.

5.3. Assigned Standard Deviation

The standard deviation for proficiency assessment also called assigned standard deviation σ_{pt} is the second key measure for the evaluation of PTs. Due to the highly variable scopes of PTs it is not feasible to fix a single method for assigning σ_{pt}. Rather the determination of σ_{pt} is up to expert opinion, taking into account factors such as the concentration of the residues and also legal requirements. The standard deviation for proficiency assessment may for example be identified as the (robust) reproducibility standard deviation, the HORWITZ standard deviation, or the THOMPSON standard deviation. The PT provider will include information on the mode for assigning σ_{pt} in the final report.

5.4. Uncertainty of the Assigned Value

There are a number of possibilities to calculate the standard uncertainty of the assigned value. For uncertainties extrapolated from PT data, usually the following formula is used:

$$u(x_{pt}) = 1.25 \frac{\sigma_R}{\sqrt{p}} \qquad \textit{with } p = \textit{number of participants}$$

However, it is also possible to sum all of the contributing uncertainties arising for example from homogeneity and stability. It is only mandatory to include information on the uncertainty of the assigned value in the final report on results if this value has an influence on the overall proficiency assessment of the participants.

5.5. Additional Statistical Parameters

Additional statistical parameters may be calculated and presented in the PT report but are only given for information purposes and are not used for the participants' proficiency assessment.

5.6. Standardised Deviation of Test Results

In order to assess the participants' proficiency it is necessary to standardise the deviation of the results from the assigned value x_{pt}. EN ISO/IEC 17043:2010 and ISO 13528:2015 state that z-scores for the quantitative results of laboratories may be calculated according to the following equation:

$$z = \frac{x - x_{pt}}{\sigma_{pt}}$$

with

x	*laboratory result*
x_{pt}	*assigned value*
σ_{pt}	*assigned standard deviation*

The z-score determination has the advantage of providing a standard value allowing to compare the results both within one proficiency test as well as between different proficiency tests, irrespective of the concentration of the analytes.

If the data are normally distributed and the reproducibility standard deviation σ_R is used as assigned standard deviation σ_{pt} the probability of the absolute value z not exceeding 2 is approximately 95 %.

Therefore, it is sensible to establish the value 2 as a quality criterion for the measurements. However, the PT provider may decide to use a different method for assigning the standard deviation for proficiency assessment σ_{pt} which may then represent a more stringent measure of performance. This can be sensible for certain residues. In that case, it is possible that more than 5 % of the values are above a z-score of 2.

Particularly in the range of the MRL value, there are high demands on the analysis of MRL substances regarding a correct quantification. Thus, the deviation from the assigned value should normally not exceed the single standard deviation for proficiency assessment which corresponds to z-scores ≤1. For banned and non-authorised compounds the identification and confirmation is more important than their accurate quantification. Therefore |z|<2 is established as a quality criterion.

Assuming a "well-behaved analytical system", EN ISO/IEC 17043:2010 suggests the following classification:

	|z|	≤2	satisfactory
2<	|z|	≤3	questionable
	|z|	>3	unsatisfactory

For MRL compounds in EURL PTs this is modified to give:

	|z|	≤1	satisfactory
1<	|z|	≤2	questionable
	|z|	>2	unsatisfactory

If the homogeneity assessment indicates that for certain analytes the standard deviation between samples is larger than anticipated, the influence of the inhomogeneity on the participant results may affect the performance assessment. To account for this, the uncertainty arising from the insufficient homogeneity can be included in the calculation of the z'-score, a modified version of the z-score:

$$z' = \frac{x - x_{pt}}{\sqrt{\left(\sigma_{pt}^2 + u^2(x_{pt})\right)}}$$

with

x	*laboratory result*
x_{pt}	*assigned value*
σ_{pt}	*target standard deviation*
$u^2(x_{pt})$	*uncertainty of the assigned value*

If the assigned value is close to zero, the confidence band around x_{pt} is asymmetrical because analytical results cannot assume negative values. In this case it may be necessary to modify the z-score to account for unwanted bias in participant performance. The method for the determination of these so-called z_U-scores is described in detail in DIN 38402-45:2014. Ultimately, the decision to use z-, z'-, or z_U-scores is made by the responsible PT provider.

A different score, which can provide useful information on the plausibility of the laboratory's measurement uncertainty, is the ζ-score:

$$\zeta = \frac{x - x_{pt}}{\sqrt{\left(u^2(x) + u^2(x_{pt})\right)}}$$

with

x	*laboratory result*
x_{pt}	*assigned value*
$u^2(x)$	*uncertainty of the participant result*
$u^2(x_{pt})$	*uncertainty of the assigned value,*

for consensus mean $u(x_{pt}) = 1.25\frac{\sigma_R}{\sqrt{p}}$ *with* $p =$ *number of participants*

The ζ-score provides information on the plausibility of the measurement uncertainty associated with an analytical result compared to the measurement uncertainty associated with the assigned value. The EURLs encourage the submission of measurement uncertainties but will only provide the ζ-score for information purposes.

The classification of z'-, z_U-, and ζ-scores is equivalent to that of the z-scores.

5.7. False Positive and False Negative Results

With respect to the general tasks of official residue control, where false negative and false positive results are a major problem regarding the trust of consumers in the system of consumer protection, all false negative and false positive results are included in the laboratory assessment. This also includes false qualitative results given in the context of a quantitative PT.

A false positive result is defined as a residue reported by a PT participant that

- is considered unlikely due to the production of the proficiency testing samples
- has not been detected by the responsible EURL or trusted expert laboratories in repeated analysis
- has not been reported by the majority of PT participants.

The ultimate decision on whether or not a reported result is considered false positive is to be made by the responsible EURL.

A false negative result is defined as a residue which has not been reported by a participant even though it has been detected by the PT organiser in repeated analysis of the sample material. False negative results include both residues which are part of the participant's method scope (not found – n. f.) as well as residues which are not but are relevant in the context of the PT (not analysed – n. a.). Should a participant's decision limit CCα exceed the concentration of the respective residue in the sample it is up to the responsible EURL to judge whether or not this is considered a false negative result, taking into account current legal limits and MMPR.

5.8. EURL Point Score System

The EURLs use a scoring system for evaluation of the participants' performance which assigns a point score based on z-scores[1] (confirmatory methods) as well as for false results. If a laboratory submits semi-quantitative results of the form "</>-value" these are checked for plausibility. In case these values are not plausible, they are treated as false negative / false positive results. Should they be plausible, they are listed in the overview of results but no z-score is derived for the residue in question.

For confirmatory methods the performance is assessed using the following point score system:

[1] And z'-/z_U-scores, respectively

Table 1: Point score system used for MRL compounds determined using a confirmation method.

Result	*Interpretation of result*	*Points*
|z-score|>2	qualitative detection performed, quantification considerably outside tolerance limits	0.5
|z-score|>1 and ≤2	qualitative detection performed, quantification outside tolerance limits	1.0
|z-score|≤1	qualitative detection performed, quantification within tolerance limits	1.5
false positive	one to x false positive results	-1.0
	more than x false positive results	-2.0
false negative		0
plausible qualitative result		0

Table 2: Point score system used for banned/non-authorised compounds determined using a confirmation method.

Result	*Interpretation of result*	*Points*
|z-score|>2	qualitative detection performed, quantification outside tolerance limits	1.0
|z-score| ≤2	qualitative detection performed, quantification within tolerance limits	1.5
false positive	one to x false positive results	-1.0
	more than x false positive results	-2.0
false negative		0
plausible qualitative result		0

Note that the same scoring system is used if z'- or z_U-scores are calculated instead of z-score. The tolerated amount x of false positive results may vary for different PTs. It is determined by the responsible EURL and communicated to the participants.

To assess participant performance for screening methods, the following point score system is applied:

Table 3: Point score system used for compounds determined using screening methods.

Result	*Interpretation of result*	*Points*
identified	qualitative detection performed	1.5
false positive	more than x % false positive results	- y
false negative		0

The responsible EURL may decide to subtract a number of points y from the participant result if the false positive rate is larger than a critical percentage x. This modification shall be communicated to the participants.

5.9. Overall Evaluation of Proficiency

The proficiency test is considered passed, if the score a participant achieves, exceeds a predetermined percentage of the maximum possible score. Usually this would be 65 % of the maximum possible score but it can be adapted to accommodate for specific requirements of a PT. In any case, the conditions required for passing will be included in the final report on results. Should the required percentage of the total score lead to point score values which are not a multiple of 0.5 points, the minimum required number of points for passing the PT is rounded to the next lowest appropriate score. For example: In a PT with five minimum required and/or recommended analytes that maximum score would be 7.5. Theoretically the minimum passing score would be 4.9, which is then rounded to 4.5.

The overall proficiency assessment is carried out in a two-step approach. In the first step, the maximum attainable number of points is calculated considering:

- regulatory aspects (e. g. MRLs, RPAs)
- the requirements set by the EURL (e. g. required analytes)
- the number of participants having provided acceptable results for the respective parameters

The calculated maximum attainable number of points allows the laboratories to compare their performance to a relevant international benchmark. For this first evaluation step all minimum required and recommended compounds contained in the samples are taken into account for the overall proficiency assessment irrespective of the inclusion of the analytes in the participant's method scope.

For all those participants who do not achieve a sufficient number of points in this first assessment round, a second assessment step is carried out. For this round only the residues actually included in the participants' methods are taken into account for the overall point score. This means that a participant who analyses only some of the minimum required and recommended residues and shows good results for these may still pass the PT but receives the feedback that they need to extend their analytical methods.

If the responsible EURL allowed the submission of results obtained using screening as well as confirmation methods, a separate evaluation for the screening and the confirmation part of the PT will be carried out following the procedure described above.

5.10. Multi-year Evaluation

In order to facilitate the identification of persisting problems for certain NRLs, a colour-coded multi-year overview is generated from the PT data. The multi-year overview is not included in the PT report but may be discussed during workshops and submitted to the Commission upon request.

Colour	*Score*	*z-scores*	*False Positives*	*False Negatives*
dark green	perfect score	all z-scores within specification	no false positive results	no false negative results
green	passed PT in 1st assessment round	majority of z-scores within specifications		acceptable amount of false negative results
yellow	passed PT in 2nd assessment round	sufficient percentage of z-scores within specifications		not all minimum required and recommended substances included in the method scope
orange	PT failed	insufficient percentage of z-scores within specifications		
red	registered for participation but did not submit results			
black	did not participate			

6. Follow-up Measures and Cooperation

The organisation of comparative testing or proficiency testing (PTs) within the NRL network is a key responsibility for EURLs. Proficiency tests for the determination of veterinary drug residues in biological matrices originating from food producing animals are organised by the responsible EURLs on a yearly base. Participation in the PTs is obligatory for the NRLs. In case of an underperformance of the NRL, it is the task of the EURL to initiate a follow up action according to the Commission protocol[2]. In the following it is described how the results of a PT are presented and communicated to the Commission as well as how underperformance in proficiency tests and a lack of collaboration on the part of the NRLs is managed.

Results of the Proficiency Test

The performance of the participants in the PT is evaluated and described in the corresponding PT report, which is accessible on the restricted webpages of the EURLs. Upon request, the EURLs provide the Commission with the coding of each laboratory.

In the case of underperformance of NRLs, the following procedure will be applied by the responsible EURL. Follow up on underperformance of official control laboratories is within the responsibility of the respective NRL.

Results of the Proficiency Test and Underperformance

The responsible EURL informs the NRLs of the publication of the PT report. In case of underperformance, the EURL will ask the NRL for feedback on the possible causes for the observed deviations and follow up action for improvement if applicable.

Underperformance includes:

a) unable to participate
b) registered for participation but did not submit results
c) failed the proficiency test
d) passed the proficiency test in the second assessment round (individual assessment)

For c) and d) a follow-up on questionable and unsatisfactory results ($|z|\geq 2$ for both MRL compounds and prohibited / non-authorised compounds) as well as false negative and false positive results is expected. In certain cases (e. g. high number of false positive results) the EURL may ask participants to provide follow-up information even though they passed the proficiency test in the first assessment round.

The feedback from the NRL must include at least a comprehensible analysis of the cause of the deviation and the proposed corrective actions and means of verification, including a proposed timeline. The NRL response to the report shall be transmitted by e-mail to the EURL usually within one month after the request for information. Upon acknowledgment of reception of the explanation, the responsible EURL shall decide if the case can be closed or if further corrective actions need to be taken. The decision will be notified to the NRL. During this process, strict confidentiality will be kept between EURL and NRL.

In case further action is needed the responsible EURL shall inform the NRL on additional follow-up measures. If deemed necessary by the responsible EURL a re-assessment of the NRL performance shall be realised after the completion of the corrective action. Follow-up measures include but are not limited to:

[2] European Commission, Protocol for management of underperformance in comparative testing and/or lack of collaboration of National Reference Laboratories (NRLs) with Community Reference Laboratories (CRLs) activities.
https://eurlcefas.org/media/4149/protocol_for_management_of_underperformance_in_comparative_testing_and_or_lack_of_collaboration_of_national_reference_laboratories.pdf

- analysis of a new PT sample set
- analysis of a different reference sample
- analysis of standard solutions
- participation in a commercial PT
- sample exchange with the EURL
- implementation of EURL test method
- participation in EURL-organised collaborative trial
- supporting visit by the EURL
- individual training

If an on-site training is required, a dedicated mission report shall be composed by the responsible EURL including an agenda of all further necessary actions and a timeline. The NRL shall inform the responsible EURL on the completion of the individual steps according to the timeline. The mission report shall be transmitted to the NRL as well as to the Commission for information.

In the case of repeated underperformance, or if corrective actions continue to result in an underperforming situation, or if the NRL does not fully collaborate to solve the requirements, the responsible EURL will officially inform the Commission and will transmit the dossier to DG SANTÉ. The Commission shall inform the competent authority of the Member State of the NRL and require that appropriate actions be taken.