JRC Scientific and Technical Reports

Guidelines for performance criteria and validation procedures of analytical methods used in controls of food contact materials

Stefanka Bratinova, Barbara Raffael, Catherine Simoneau

Community Reference Laboratory
CRL
Food Contact Materials

EU Network of National Reference Laboratories
NRL
Food Contact Materials

JRC EUROPEAN COMMISSION

ihcp Institute for Health and Consumer Protection

EUR 24105 EN - 1st edition 2009

The mission of the IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.

The IHCP carries out research to improve the understanding of potential health risks posed by chemical, physical and biological agents from various sources to which consumers are exposed.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: DG JRC, IHCP TP 260, I-21020 Ispra
E-mail: Catherine.simoneau@jrc.ec.europa.eu
Tel.: 39.0332.785889
Fax: 39.0332.785707

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 53034

EUR 24105 EN
ISBN 978-92-79-14483-7
ISSN 1018-5593
DOI 10.2788/49046

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Executive Summary

Test methods for materials and articles in contact with foodstuffs are required to determine the concentration of residues of monomers in the materials themselves or to determine the concentration of individual or groups of substances in food (or food simulants) which have migrated from the food contact materials.

The Community Reference Laboratory and National Reference Laboratories for food contact materials (FCM) prepared the present Guidelines to illustrate the required performance criteria for the analytical methods applied in the laboratories for FCM and provide procedures for method validation in order to estimate their performance characteristics. The scope of these guidelines is to provide rules for the performance of the analytical methods to be used in the verification of compliance with the migration limits defined in Directive 2002/72/EC,as amended, and in accordance with Directive 82/711/EEC, as amended, and others defined in the European legislation, in order to ensure the quality and comparability of the analytical results.

The document presents 4 approaches, according to the different purpose of performance assessment.

These guidelines are intended as a dynamic document and they will evolve and expand into further editions. This is the first edition. These guidelines have been endorsed by the European Union official Network of National Reference Laboratories and approved by the EU Commission competent service DG SANCO.

This work also highlights an important deliverable for the Network of NRLs. In particular, the members of the task force “Method Performance” that have dedicated time and effort to provide input into the development of these guidelines. They are gratefully acknowledged here for their contribution: NRL-BE (Fabien Bolle, Tina n’Goy), NRL-DE (Oliver Kappenstein), NRL-DK (Jens Petersen), NRL-ES (Juana Bustos), NRL-FR1 (Patrick Sauvegrain), NRL-EL (Timokleia Togkalidou), NRL-IT (Maria Rosaria Milana), NRL-NL (Durk Schakel, Dita Kalsbeek-van Wijk), NRL-PL (Kazimiera Cwiek-Ludwicka), NRL-SI (Viviana Golja), NRL-UK (Emma Bradley). Special thanks are extended to Emma Bradley for her contribution to the editing of the document.

Acknowledgement:
This work was performed thanks to the CRL mandate:
SANCO/2007/FOODSAFETY/0043-Contact Materials
SANCO/2008/FOODSAFETY/0047-Contact Materials
SANCO/2008/FOODSAFETY/055-Contact Materials

Table of Contents

1 INTRODUCTION

The European Framework Regulation (EC) No. 1935/2004 [29] is the basic Community legislation that covers all food contact materials and articles. It empowers the European Commission to set requirements for specific materials. Specific requirements for such materials can include limits on the overall migration and on the specific migration of certain constituents or groups of constituents into foodstuffs. These limits have been defined for some substances in plastic materials and articles.

Test methods for materials and articles in contact with foodstuffs are required to determine the concentration of residues of monomers in the materials themselves or to determine the concentration of individual or groups of substances in food (food simulants) which have migrated from the food contact materials or to determine overall migration from food contact materials.

The determination of migration from materials and articles intended to come into contact with foodstuffs is quite unlike any other measurement tasks in ensuring food safety and quality. Reliable measurements depend upon more than simply having validated analytical methods for measuring chemical concentrations in foods. The Directives allows, as an alternative to the analysis of the foodstuff itself, migration testing to be carried out with food simulants applied under conditions which simulate actual use of the material or article with food. This introduces additional potential sources of variability in the final migration value.

It is necessary to ensure the quality and comparability of the analytical results generated by laboratories for enforcement purposes, for compliance purposes, and for the creation of data for risk assessment purposes. This should be achieved by using quality assurance systems and specifically by applying methods that have been validated according to common procedures and that meet defined performance criteria, and by ensuring traceability to common standards or standards that are commonly agreed upon.

Commission Regulation (EC) No. 882/2004 [30] on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules, which also incorporates food contact materials, requires official control laboratories to be accredited according to EN ISO/IEC 17025 [21]. Moreover, approved laboratories must prove their competence by regular and successful participation in adequate proficiency testing schemes recognised or organised by the national or Community reference laboratories.

A network of Community Reference Laboratory and National Reference Laboratories for food contact materials (FCM) operates under Commission Regulation (EC) No. 882/2004 [30] to enhance coordination.

In the field of materials and articles in contact with food numerous chemicals are used in the manufacturing processes and it is not possible to prepare standard test methods for all. Therefore the concept of routine methods and reference methods should be superseded by a criteria approach, in which performance criteria and procedures for the validation of screening and confirmatory methods are defined.

It is necessary to determine common criteria for the interpretation of test results of official control laboratories in order to ensure a harmonised implementation of Commission Regulation (EC) No. 882/2004 [30].

2 SCOPE OF THIS DOCUMENT

The scope of these guidelines is to provide rules for the performance of the analytical methods to be used in the verification of compliance with the migration limits defined in Directive 2002/72/EC [22], as amended, and in accordance with Directive 82/711/EEC [23], as amended, and 85/572/EEC [24], as amended, and the further provisions set out in Annex 1 of Commission Directive 2002/72/EC [22].

The verification of compliance with the migration limits is made using methods that:

(a) are documented in test instructions, preferably according to ISO 78-2 [20];

(b) comply with the performance criteria defined in these guidelines;

(c) have been validated according to the procedures described in these guidelines.

The quality of the results of the analysis of samples for verification of the compliance with the migration limits should be ensured according to Chapter 5.9 of ISO 17025 [21].

3 GLOSSARY - DEFINITIONS

Abbreviations	Explanation
Accuracy	The closeness of agreement between a test result and the accepted reference value. It is determined by determining trueness and precision.
Analyte	The substance, and any derivatives emerging during its analysis, that has to be detected, identified and/or quantified.
Bias	The difference between the expectation of the test result and an accepted reference value
Calibration standard	A device for measurements that represents the quantity of substance of interest in a way that ties its value to a reference base
Certified reference material (CRM)	A material that has had a specified analyte content (or for food contact materials a migration value) assigned to it
Collaborative study	Analysing the same sample by the same method to determine the performance characteristics of the method. The study covers random measurement error and laboratory bias
Compositional limit (QM)	The maximum permitted amount of the named substance in the material or article.
Confirmatory method	Methods that provide full or complementary information enabling the substance to be unequivocally identified and if necessary quantified at the level of interest.
Food Simulant	A medium intended to simulate (‘mimic’ or ‘model’) the essential characteristics of a foodstuff.
Fortified sample material	A sample enriched with a known amount of the analyte to be detected
Interlaboratory study (comparison)	Organisation, performance and evaluation of tests on the same sample by two or more laboratories in accordance with predetermined conditions to determine testing performance. According to the purpose the study can be classified as collaborative study or proficiency study.
Internal standard (IS)	A substance not contained in the sample with physical-chemical properties as similar as possible to those of the analyte that has to be identified and which is added to each sample as well as to each calibration standard.
Level of interest	The concentration of a substance or analyte in a sample that is significant to determine its compliance with legislation.
Overall migration	The mass of material transferred to the food simulant or test media as determined by the relevant test method.
Performance characteristic	Functional quality that can be attributed to an analytical method. This may be for instance specificity, accuracy, trueness, precision, repeatability, reproducibility, recovery, detection capability and ruggedness.
Performance criteria	Requirements for a performance characteristic according to which it can be judged that the analytical method is fit for the purpose and generates reliable results.
Precision	The closeness of agreement between independent test results obtained under stipulated (predetermined) conditions. The measure of precision usually is expressed in terms of imprecision and computed as standard deviation of the test result. Less precision is determined by a larger standard deviation.
Proficiency study	Analysing the same sample allowing laboratories to choose their own methods, provided these methods are used under routine conditions. The study has to be performed according to ISO guide 43-1 [31] and 43-2 [32] and can be used to assess the reproducibility of methods.

Range (working or measuring)	Means a set of values for which a measure is intended to lay within specified uncertainty limits
Recovery	The percentage of the true concentration of a substance recovered during the analytical procedure. It is determined during validation, if no certified reference material is available.
Reference material	A material of which one or several properties have been confirmed by a validated method, so that it can be used to calibrate an apparatus or to verify a method of measurement.
Repeatability conditions	Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment and short interval of time.
Repeatability	Precision under repeatability conditions (r) - the value below which the absolute difference between 2 single test results obtained under repeatability conditions, may be expected to lie within a specific probability (typically 95 %) and hence r = 2,8 x sr.
Reproducibility conditions	Conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.
Reproducibility	Precision under reproducibility conditions (R) - the value below which the absolute difference between 2 single test results obtained under reproducibility conditions, may be expected to lie within a specific probability (typically 95 %) and hence R = 2,8 x sR.
Residual content	The mass of the substance present in the final material or article.
Ruggedness	The susceptibility of an analytical method to changes in experimental conditions which can be expressed as a list of the sample materials, analytes, storage conditions, environmental and/or sample preparation conditions under which the method can be applied as presented or with specified minor modifications. For all experimental conditions which could in practice be subject to fluctuation (e.g. stability of reagents, composition of the sample, pH, temperature) any variations which could affect the analytical result should be indicated.
Sample blank determination	The complete analytical procedure applied to a test portion taken from a sample from which the analyte is absent.
Screening method	Methods that are used to detect the presence of a substance or class of substances at the level of interest. These methods have the capability for a high sample throughput and are used to sift large numbers of samples for potential non-compliant results. They are specifically designed to avoid false compliant results.
Single laboratory study (in-house validation)	An analytical study involving a single laboratory using one method to analyse the same or different test materials under different conditions over justified long time intervals.
SML(T)	The maximum permitted level of a group of named substances migrating from the final material or article into food or food simulants, expressed as total of chemical moiety or substance(s) indicated.
Specific Migration	The mass of the substance transferred to the food/simulant as determined in the test method.
Specific Migration Limit (SML)	The maximum permitted level of a named substance migrating from the final material or article into food or food simulants.
Specificity	Ability of a method to distinguish between the analyte being measured and other substances. This characteristic is predominantly a function of the measuring technique described, but can vary according to class of compound or matrix.
Standard addition	A procedure in which the test sample is divided in two (or more) test portions. One portion is analysed as such and known amounts of the standard analyte are added to the other test portions before analysis. The amount of the standard analyte added has to be between two and five times the estimated amount of the analyte in the sample. This procedure is designed to determine the content of an analyte in a sample, taking account of the recovery of the analytical procedure.
Trueness	The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. Trueness is usually expressed as bias.
Validation	The confirmation by examination and the provision of effective evidence that the particular requirements of a specific intended use are fulfilled.
Within-laboratory reproducibility	Precision obtained in the same laboratory under stipulated (predetermined) conditions (concerning e.g. method, test materials, operators and environment) over justified long time intervals.

4 METHOD VALIDATION PLAN

4.1 Choice of validation scheme

The detailed design and the correct execution of method validation studies should, as far as possible, provide a realistic assessment of the number and range of effects operating during normal use of the method, as well as covering the working

concentration range(s) and sample types that fall within the scope of the method.

The applicable working concentration range is an important part of the validation or verification of the analytical method. It will often save both time and effort to choose the area of application on the basis of laboratory or Regulatory needs, instead of validating the whole range of possibilities.

It is also important that the validation and verification report describes which sample matrices have been used. In fact, the performance of some methods may be also dependent on the matrix. In such cases, it is very important to ensure that the entire stated area of application is included in the validation.

Validation or verification of an already validated method must always be performed before it is used in the laboratory for official control purposes. This work must be repeated partly or fully if the result of the first validation makes it necessary to modify the method.

In the following paragraphs four different approaches are presented:

4.1.1 "Full" single laboratory validation protocol applicable to the field of food contact materials and articles.

Full validation should be performed for newly developed methods or methods published in scientific literature, but without important performance characteristics, that are to be used for standardisation purposes or in public control.

"Full validation" of a method means thorough examination and determination of the performance characteristics of the method. Validation should demonstrate that the analytical method complies with the established criteria applicable for the relevant performance characteristics.

A single-laboratory validation cannot be considered a real full validation and some of the parameters assessed have a consistent value only for the laboratory that performed the validation (more details will be provided in the specific chapters).

4.1.2 "Standard level" of single laboratory validation applicable to the field of food contact materials and articles

This scheme represents the conditions specifically developed and agreed by the official control laboratories for FCM which should be applied as a working standard for use in the field of FCM. This validation scheme represents the minimum requirements to establish non-compliance of a material or article intended for food contact.

Experienced laboratories, who have already implemented more sophisticated procedures, compliant with consolidated standards of validation, can continue to use them provided they respect at least the minimum requirements set by the agreed level described in these guidelines.

4.1.3 "Basic level" of single laboratory validation applicable to the field of food contact material and articles

This scheme represents the base level that must be met by all laboratories. This level does not fulfil all of the legal and official requirements, but it is considered as

the starting point from which a harmonised level of control (and results) in all European countries can be achieved. From this starting point the laboratories will have to update and improve their procedures and achieve the level foreseen for the “Standard level” validation by January 2011.

This “Basic level” validation may also be used for emergency or occasional cases and/or for a reduced number of samples. In such cases this approach may be used beyond 2011.

Experienced laboratories, who have already implemented more sophisticated procedures, compliant with consolidated standards of validation, can continue to use them provided they respect at least the minimum requirements set by the agreed level described in these guidelines.

4.1.4 Method verification

Method verification is the examination of a laboratory's ability to perform the analysis in accordance with the method parameters established in the validation of the method. This scheme should be followed to prior to the use of an already validated method (by another laboratory) or by the same laboratory but that has not used the method for a defined period of time or when the method is used regularly but the method performance has not been checked for a defined period of time.

The extent of the verification performed locally in each laboratory depends on how thoroughly the method has been validated: a thorough validation simplifies the internal verification. It should be emphasised that some methods, although issued by standardisation bodies, have not been validated through collaborative studies and have to be thus validated and not only verified.

Verification should be performed every year if the method is used regularly.

Figure 1-4 show the respective flow charts for the four validation and verification processes described above.

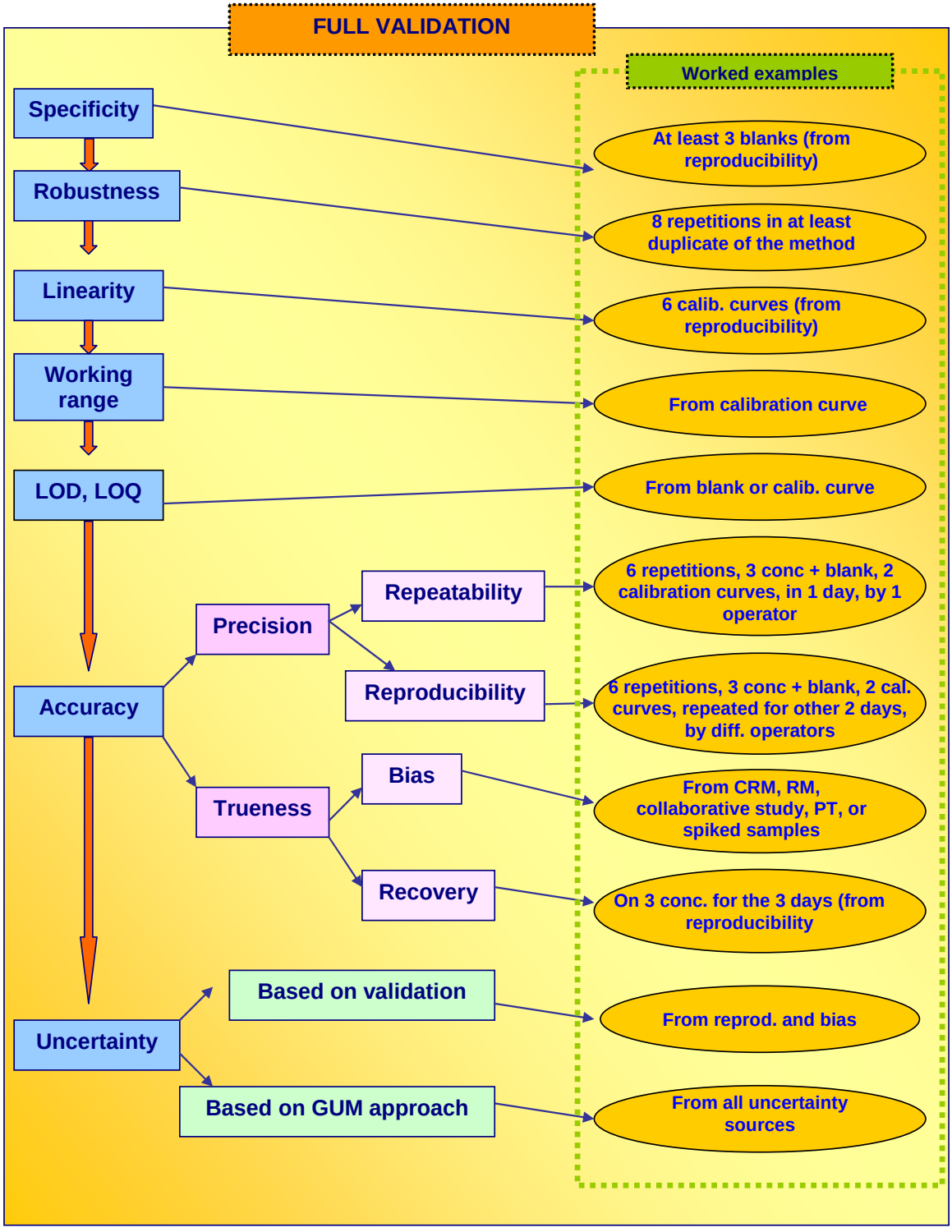


Figure 1: Flow chart of a full validation scheme for FCM

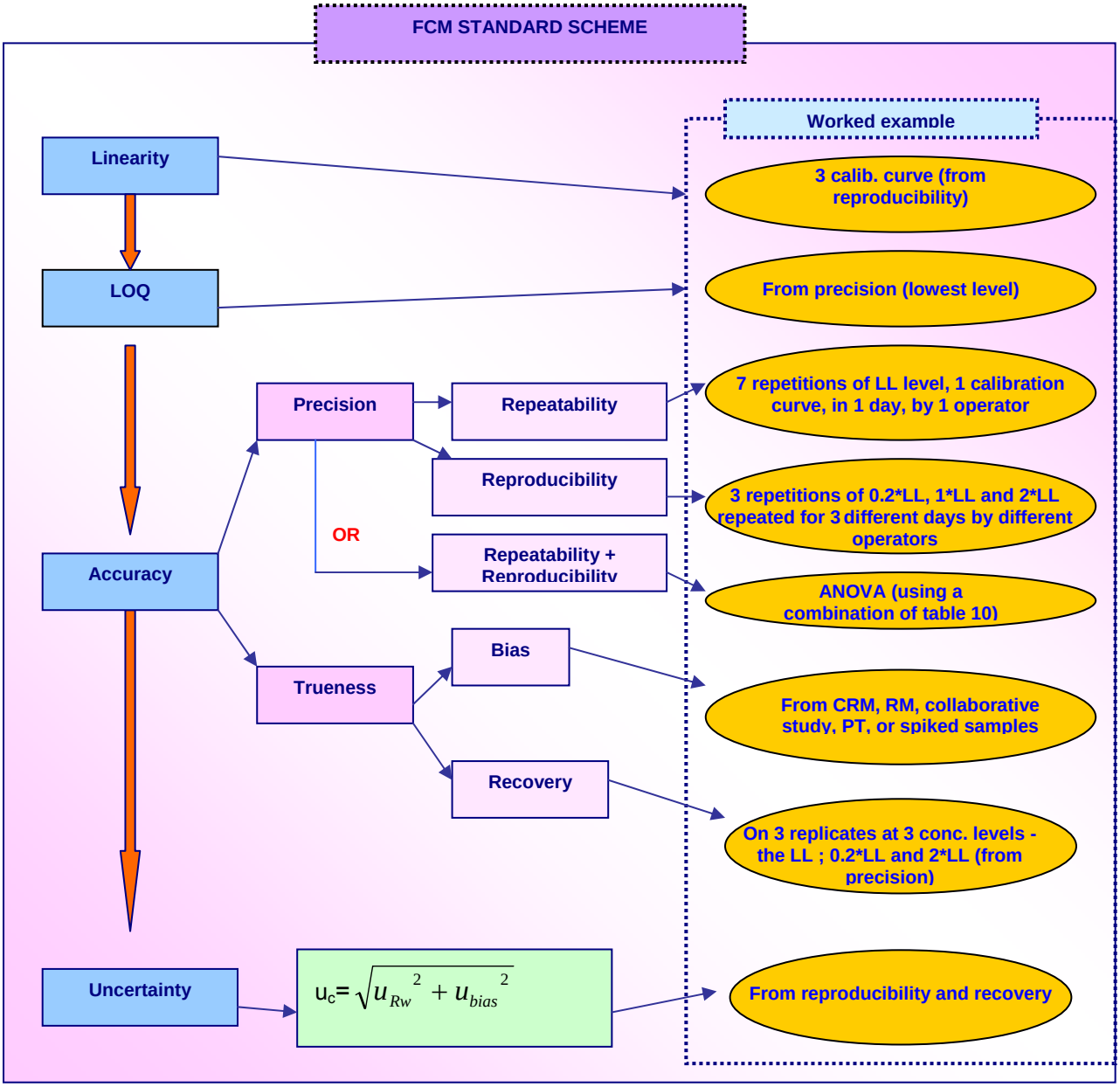


Figure 2: Flow chart of the effective standard working validation scheme for FCM

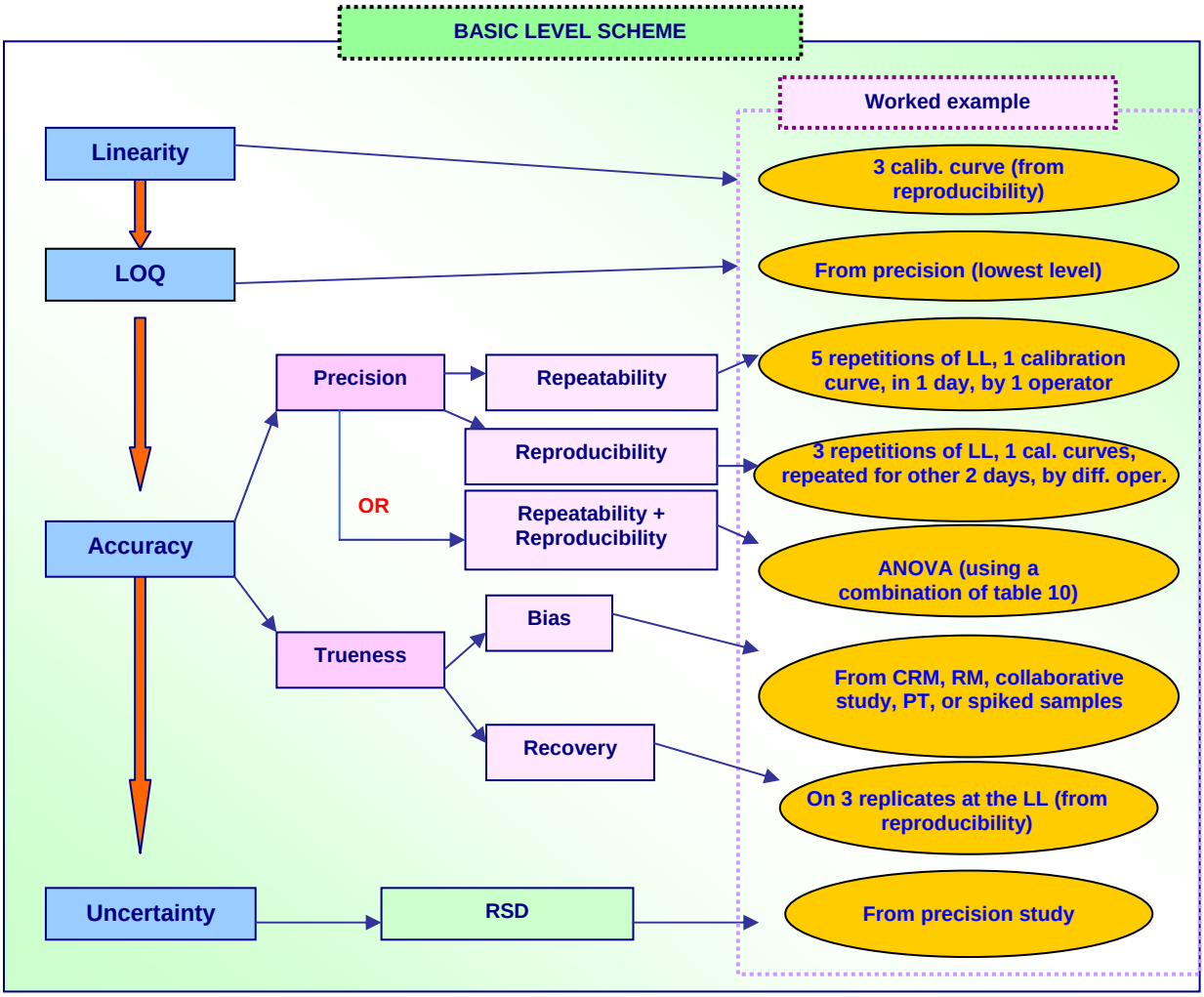


Figure 3: Flow chart for bottom level validation process

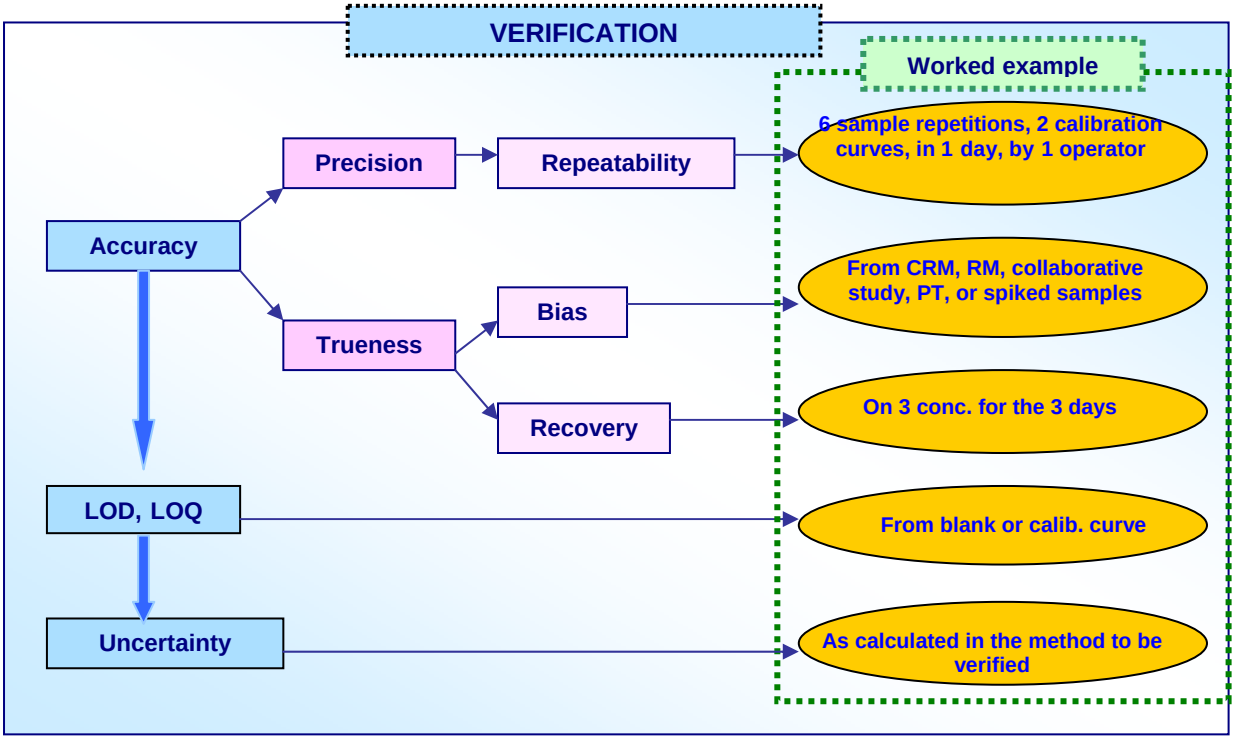


Figure 4: Flow chart for a verification process

5 “FULL” SINGLE-LABORATORY METHOD VALIDATION PROCEDURE

In the validation of test methods in the field of FCM the following cases could be encountered:

- An overall migration test for the determination of the total quantity of all non-volatile substances that have migrated from the FCM test specimen to a food simulant
- A specific migration test for the determination of the substances in food simulants that have migrated from a food contact material or article
- The determination of the substance(s) in food, that have migrated from a food contact material or article
- The determination of residues of monomers in food contact materials and articles intended to come into contact with foodstuffs.

In an overall migration test the total quantity of all of the substances that have migrated from the test specimen of a food contact material or article to a food simulant is determined gravimetrically using the procedures outlined in European Commission Directive 2002/72/EC [22], as amended, and the detailed methods in the EN 1186 series of standards.

In a specific migration test the quantity of an individual substance (monomer, additive, etc.) or group of substances is determined in a food simulant following the exposure of a food contact specimen to the food simulant for the prescribed period of time at the prescribed temperature using an appropriate analytical method.

Both of these migration tests consist of two parts - the exposure of the test specimen to the food simulant (or food for specific migration) and the analytical determination part. The performance characteristics of both a specific migration test and an overall migration test are a combination of factors from the migration part and the determination part of the test. The outcome of the exposure part (and therefore the test result) is furthermore dependent on the material tested e.g. degree of homogeneity and interaction with the food or food simulant and the test conditions applied.

The performance characteristics for the determination of substances in food (when the material is already in contact with the food) and the determination of residues of monomers in food contact materials are just those of the determination part (analytical method).

The methods described allow compliance with the following legislative limits (LL) to be demonstrated:

- QM= Maximum permitted quantity of the ‘residual’ substance in the material or article;
- QM(T)= Maximum permitted quantity of the ‘residual’ substance in the material or article expressed as total of moiety or substance(s) indicated.
- QMA= Maximum permitted quantity of the ‘residual’ substance in the finished material or article expressed as mg per 6 dm^2 of the surface in contact with foodstuffs.

- QMA(T) = Maximum permitted quantity of the 'residual' substance in the material or article expressed as mg of total of moiety or substance(s) indicated per 6 dm² of the surface in contact with foodstuffs.
- SML = Specific migration limit in food or in food simulant, unless it is specified otherwise.
- SML(T) = Specific migration limit in food or in food simulant expressed as total of moiety or substance(s) indicated.
- OML= Overall migration limit is the maximum permitted total amount of substances that may be released from a material or article into food or food simulant.

In the following paragraphs the term Legislative Limit (LL) will be used for all the above type of limits

5.1 Performance characteristics of the migration part

Performance characteristics of the migration part depend on the variability in the migration contact stage.

The time/temperature conditions are specified in Directive 82/711/EEC [23], as amended by Directives 93/8/EEC [33] and 97/48/EC [34] for packaging materials, but not for kitchen articles kitchenware and cookware, i.e. articles placed in contact with food in the consumer's daily use. In order to minimise possible variability in the migration contact stage it is highly important that harmonised test conditions are applied (as expressed in the "Guidelines on testing conditions for materials and articles in contact with foodstuffs", EU Report EUR 23814 EN 2009).

As it is well described in EN 1186 even for situation where there are well defined test/temperature conditions the tolerances placed on the exposure period and on the test temperature introduce a source of variability in the final migration value, for both overall and specific migration.

The standard exposure times range from 30 minutes to 10 days, and the standard exposure temperatures range from 5ºC to 175ºC. It is well known that migration is a diffusion process which increases with increasing temperature and time. The tolerances for the contact times and temperatures defined in the legislation are reported in the tables below:

Contact time (hours)	Tolerance +/- (min)
0.5	+1/-0
1	+2/-0
2	+5/-0
4	+10/-0
24	+30/-0
240	+300/-0

Contact temperature (ºC)	Tolerance +/- (ºC)
5	1
20	1
40	1
70	2
100	3
121	3
130	5
150	5
175	5

Table 1a and b: Time and temperature tolerances

According to the CEN TR 15330 tests performed with short times and high temperatures will be expected to have a lower precision than tests performed with

long times and low temperatures, due to difficulties in controlling/replicating exposure conditions, particularly temperature. Most high temperature tests are performed with exposure times of less than 24 hours and special care is needed to control the heating-up and cooling-down stages of the exposure of the specimen to the hot simulant.

Another potential source of variability in the migration contact stage is the conventional assumption that the fat simulant D - olive oil, and its substitute solutions sunflower oil and the synthetic triglyceride HB307 - are equivalent. This is unlikely to be strictly true since these oils differ both in their chemical composition (e.g. chain-length) and physical composition (e.g. viscosity). It can be expected that a laboratory using one of the choices of simulant D will, under some circumstances, obtain a different test result, for SM or OM, than a laboratory using one of the other choices of simulant D.

As a result of all mentioned above in the validation of test methods in the FCM field the performance characteristics of one migration method should be determined at strictly controlled migration conditions.

Accuracy and its components precision and trueness, as well as uncertainty as a general term assessing accuracy are the performance characteristics which could be assessed for the migration part.

Two main approaches to the estimation of uncertainty of the migration part could be described: an empirical approach and a modelling approach.

The empirical approach uses repeated migration procedures and analysis, under various conditions, to quantify the effects caused by variability of conditions to quantify uncertainty (and usually some of its component parts). Uncertainty of measurements as a common expression of the most important performance characteristics can be considered to be generated by broad sources of error. These four sources are the random errors and the systematic errors arising from the methods of both the migration and the analysis. These errors have traditionally been quantified as the migration precision, analytical precision, migration bias and the analytical bias respectively, as shown in Table 2.

	Precision (Random effects)	**Trueness (Systematic effects)**
Migration (exposure) part	Migration variability	Migration bias
Determination (Analytical) part	Analytical variability	Analytical bias

Table 2: Uncertainty contributions of one migration test method in the empirical approach

The modelling approach uses a predefined model that identifies each of the component parts of the uncertainty, making estimates of each component, and sums them in order to make an overall estimate. This can be done using specific softwares.

The contribution of the migration test to the total uncertainty could be small but most probably it represents the dominant effect. Therefore to reduce the total uncertainty of the method the exposure phase should be well controlled in order to achieve fitness for purpose.

Such Guidelines can never be fully comprehensive, and although there are details of some of the statistical techniques employed and references to more detailed advice,

there will often be a need for expert judgment for more complex situations.

If errors from these four sources (migration variability, migration bias, analytical variability and analytical bias) could be quantified, separately or in combinations, it is possible to estimate the uncertainty of the measurements that these migration methods could produce.

Methods for the estimation of three of the four errors are established. Analytical precision and analytical bias can be estimated by one of the approaches described in the paragraph (6.7), concerning analytical part of the method (Table 2).

Migration precision can be estimated by replication of the migration procedure.

$$s^2_{migration} = s^2_{total} - s^2_{analytical}$$

where

$s_{migration}$ is the standard deviation of the migration part;

$s_{analytical}$ is standard deviation of the analytical part;

s_{total} is the standard deviation of the whole procedure (migration + analytical part).

The calculations can be made applying analysis of variance (ANOVA)

It should be taken into consideration that the precision obtained after replicate study of the migration procedure will be an estimation of the overall precision including the one coming not only from the migration part itself but in many cases also due to the inhomogeneity of the material/article tested.

	Precision (Random effects)	**Trueness (Systematic effects)**
Migration (Exposure) part	Replicate migration experiment and analysis	CRM (overall migration) Proficiency Test (PT)
Determination (Analytical) part	Replicate analyses	CRMs, PT, Recovery

Table 3: Estimation of uncertainty contributions in the empirical approach

Hence the focus of interest in validation study of the migration part will be exclusively the precision aspect. As there are no CRM for specific migration, it is not possible to assess bias in migration validation. It is though possible to assess bias of the overall migration, in this case some CRM are available.

5.2 Performance characteristics of the determination (analytical) part

The following performance characteristics are applicable to the analytical determination of migrant concentration in the food or food simulant and to the determination of the residual monomer content in a material or article.

5.2.1 Selectivity/specificity

Selectivity is the degree to which the method can quantify the target analyte in the presence of other analytes, matrices, or other potentially interfering materials. This is usually achieved by isolation of the analyte through selective solvent extraction, chromatographic or other phase separations, or by application of analyte-specific techniques such as biochemical reactions (enzymes, antibodies) or instrumentation (nuclear magnetic resonance spectroscopy, infrared or mass spectrometry).

In the validation study, at least 3 “processed” blanks have to undergo the whole procedure before being analysed. This allows any interfering substances derived from the analytical process to be identified. For chromatographic methods then there should be no peak(s) at the same retention time as the analyte of interest. In case this is not possible a limit should be defined during the validation study which guarantees that the interfering peak does not have an impact on the results.

The analysis of the processed blanks is the same as for the within-laboratory reproducibility assessment (5.2.7.1.3) i.e. at least 1 blank should be processed for each day that the method validation is performed.

Moreover, to prove that the method is selective enough, a check could be performed by spiking any potential interference substances to verify that what the method is measuring is the target substance 5.2.7.2.2.6

5.2.2 Ruggedness

The study of ruggedness is generally part of method development. If it was not studied during method development and there is the need to perform such a study, the procedure described in this chapter should be followed.

According to [8], ruggedness is defined as: “The susceptibility of an analytical method to changes in experimental conditions which can be expressed as a list of the sample materials, analytes, storage conditions, environmental and/or sample preparation conditions under which the method can be applied as presented or with specified minor modifications. For all experimental conditions which could in practice be subject to fluctuation (e.g. stability of reagents, composition of the sample, pH and temperature) any variations which could affect the analytical result should be indicated.”

In practice it means that a series of suitability parameters have to be established to ensure that the validity of the analytical method and the quality of the analytical results are maintained whenever the method is used. Thus, reasonable variations of such parameters are deliberately introduced to observe the effects and to ensure the performance of the method under different laboratory conditions.

Such parameters are “non-procedure-related factors” such as the laboratory performing the analyses, different analysts, different analytical instruments, different lots of reagents, etc.

Ruggedness represents the degree of reproducibility of the results obtained under a variety of conditions, expressed as %RSD. The International Conference for Harmonisation (ICH) addresses ruggedness as intermediate precision and reproducibility.

For in-house validation of methods, it is not possible to change these non-procedure-related parameters, so instead the analysis of the robustness of the method can used. The robustness of the method foresees the change of “procedure-related factors” (such as elution phase, quantity of elution solution used) to verify that these

small but deliberate variations in the method do not have any effect on the final results, providing an indication of the method's or procedure's suitability and reliability during normal use.

Robustness traditionally has not been considered as a validation parameter in the strictest sense because usually it is investigated during method development, once the method is at least partially optimised. Nevertheless, for the completeness of the present study it could be useful to evaluate also this parameter, to provide better and more specific information on the method use, in case a full validation of such a method will be carried on.

In Table 4 the analytical scheme (according to Youden ruggedness trial) is reported: the symbol + refers to the original parameters (A-G), while the symbol – refers to the parameter that has been changed (a-g). The scheme foresees 8 repetitions of the method, changing each time some of the seven parameters for which robustness is being assessed.

If all seven parameters are changed, an additional repetition of the method performed with the original parameters has to be carried out, to serve as a blank for reference.

It is possible to change also less than 7 parameters. In this case 8 experiments have in any case to be carried out, using the original parameters, and the experiment(s) carried out with the original parameters will represent the “blank”.

		TRIAL							
+	-	1	2	3	4	5	6	7	8
A	a	+	+	+	+	-	-	-	-
B	b	+	+	-	-	+	+	-	-
C	c	+	-	+	-	+	-	+	-
D	d	+	+	-	-	-	-	+	+
E	e	+	-	+	-	-	+	-	+
F	f	+	-	-	+	+	-	-	+
G	g	+	-	-	+	-	+	+	-

Table 4: Robustness test scheme

Worked example

Samples: A total of 16 analyses have to be performed (8 times 1 concentration in 2 replicates).
=> 16 analyses

Calibration curve: A total of 12 analyses had to be performed (5 concentrations+ blank each one analysed 2 times). If the robustness assessment in carried on in the same day of repeatability or reproducibility assessment no new calibration curve has to be prepared.
=> 12 analyses

=> 16 (+12) analyses by operator A or B

For each of the 8 experiments one concentration value will be calculated (y_1, y_2, y_3, …y_8).

These values will be combined, according to the sign shown in Table 4 and divided by half of the number of the experiments, to give a reference value for each parameter (Ei):

$E_A=(y_1+ y_2+y_3+y_4- y_5-y_6-y_7-y_8)/4$

$E_B=(y_1+ y_2-y_3-y_4+y_5+y_6-y_7-y_8)$ /4

$E_C=(y_1- y_2+y_3-y_4+y_5-y_6+y_7-y_8)$ /4

$E_D=(y_1+ y_2-y_3-y_4-y_5-y_6+y_7+y_8)$ /4

$E_E=(y_1-y_2+y_3-y_4-y_5+y_6-y_7+y_8)$ /4

$E_F=(y_1+ y_2+y_3+y_4+y_5+y_6+y_7+y_8)$ /4

$E_G=(y_1+ y_2+y_3+y_4+y_5+y_6+y_7+y_8)$ /4

The standard deviation for the effects can then be calculated:

$$SD = 2\sum\left(\frac{E_i^2}{n}\right)$$

Where:

E_i is each of the calculated effects,

n is the total number of parameters.

If the standard deviation of the results for the robustness assessment is lower or comparable to the standard deviation calculated for the reproducibility assessment (or within-laboratory reproducibility) it means that performance of the method was not influenced by the small changes in the parameters and is thus overall robust.

To assess if the method is robust compared to each variation, the influence of each variable on method performance has to be assessed applying the t-test. The experimental t values for each of the effects should be calculated using the following formula:

$$t = \frac{Ei \cdot \sqrt{n}}{SD \cdot \sqrt{2}}$$

If the calculated t-value results are lower than the t critical value for ν=n − 1 degrees of freedom (t_{crit} = 2.45, ν=6, 97.5% confidence level), then the method is robust with respect to that change in the procedure, otherwise the method should be considered sensitive for that parameter and in the procedure particular care should be taken to ensure that the method is followed exactly.

5.2.3 Calibration range, assessment of the calibration function

5.2.3.1 Calibration range

The calibration range for any method in the field of FCM should be carefully chosen with regards to the relevant legislative limits (LL) as follows:

5.2.3.1.1 Substances with specific migration limits (SMLs)

The target calibration range should contain at least five points equally distributed in the range 0.2*SML [1] up to 2*SML. When it is necessary to expand the calibration range, this should be done by including extra points above 2*SML.

The calibration range of the analytical method for fatty food simulants (simulant D) must be extended beyond 2*SML according to the reduction factor applicable to the foodstuff (up to 5 times) or the sample should be diluted prior to instrumental analysis to fit within the calibration range.

[1] **if this is not possible the lowest point in the calibration should be at least equal to the LOQ, providing that LOQ+ULOQ < SML.**

5.2.3.1.2 Substances where the SML is "Non-detectable" with a defined detection limit

The target calibration range should contain at least 5 points equally distributed in the range from 1 to 10 times the LOQ, where LOD=LOQ/2 should be less or equal to ND (the detection limit) defined in the legislation. Careful estimation and verification of the LOQ is mandatory for such substances, see 5.2.6.1.

5.2.3.1.3 Substances with Group specific migration limits (SML(T))

Substances expressed as SML(T) can fall into different categories:

5.2.3.1.3.1 Substances where the SML(T) is expressed as one equivalent analyte (e.g. as acrylic acid, as tin etc)

In this case the analytical method targets one final analyte rather than different species and therefore the requirements can be as those for a single substance SML as described above.

5.2.3.1.3.2 Substances where the SML(T) represents a known/unknown number of substances (e.g. DIDP and DINP, or BADGE and derivatives)

The target calibration range for each individual substances should contain at least 5 points equally distributed in the range from 0.2*SML(T)/n [1] up to 2*SML(T) [2] of the group (with n = number of substances included in the SML(T)). If the number of substances is not known, n should be estimated based on expertise and expectations. The upper limit does not need to be divided.

[1] if this is not possible the lowest point in the calibration should be at least equal to the LOQ of each substance, providing that Σ ($LOQ_n + U_{LOQ\,n}$) < SML(T).

[2] if this is within the linear range of the detector (e.g. determination of BADGE with fluorescence detection for which linearity is lost at high levels).

5.2.3.1.4 Substances where SML(T) = non-detectable with established detection limit and represents an unknown number of substances (e.g. functional barriers)

The target calibration range should contain at least 5 adequately distributed points in the range between 1 and 10 times LOQ;

The LOD should be the detection limit that is defined as 'not detectable' in the legislation / n, where n (the expected number of substances) should be estimated case by case based on expertise and expectations.

5.2.3.1.5 Substances with a QM or QMA

The target calibration range should contain at least five points equally distributed in the range from 0.2 * QM [1] up to 2 * QM. When it is necessary to expand the calibration range, this should be done with the inclusion of extra points at levels above 2 * QM.

1 if this is not possible the lowest point in the calibration should be at least equal to the LOQ, providing that LOQ + ULOQ) < QM.

5.2.3.1.6 Substance with QM(T) and QMA(T)

Same as for SML(T) see 5.2.3.1.3.

5.2.3.2 Basics of calibration and quantification

A calibration function is determined from values of the measurement response (y_i) at given concentrations x_i.

Depending on the type of correlation between the analyte concentration and the measurement response different mathematical or statistical tools can be used. Mathematical and statistical models themselves are subject to different assumptions. For final results not only the performance of the analytical equipment but also the appropriate choice of the mathematical approach to be used is essential.

The models should help the analyst to find a clear and reliable functional relationship for calibration and should not limit the capabilities of the analytical equipment. Therefore the selection of the calibration model should be done very carefully taking into consideration the fitness for purpose and the measurement uncertainty required.

The following calibration models should be tested for their adequateness, which means that the model that provides the lowest measurement uncertainty should be used:

- Model of linear regression [25]: it is the classical model for calibration which assumes a linear correlation between measured values (y_i) and corresponding concentrations (x_i). It has a limitation: it requires the normal distribution of the responses and homogeneity of variances over the working range. Normally this homogeneity of variances is only given in a working range of one or at maximum two orders of magnitude of concentration. This fact limits its applicability.
- Model of non-linear second order calibration functions [26]: It should be considered when the adequateness of the linearity model cannot be proved, thus after discovering a significant non-linearity. This model has the same limitations as linear regression: normal distribution of the responses and homogeneity of variances over the working range.
- Model of weighted regression: it consists on weighting each value with the reciprocal variance and it should be applied when the limitations of the other models make them not suitable for the purpose of the method. With this model a calibration over more than one order of magnitude of concentration range is possible.

Usually the calibration function in the chosen working range is a priori supposed to be linear, but this should be verified in the course of the validation study.

In the linear model the three main parameters (a, b, R^2) are defined by the equations:

$y = a+bx + \varepsilon_i$

Where:

a is the intercept;

b is the slope;

x is the concentration

y is the measurement response

ε_i is the difference between the measured value and the expected value

$\varepsilon_i = y_i - \hat{y}_i$

and

$$R^2 = \left[\frac{\sum \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum \left(x_i - \bar{x}\right)^2 \sum \left(y_i - \bar{y}\right)^2}} \right]^2$$

where

R^2 is the linear correlation coefficient

$\bar{x}$ is the mean concentration value

$\bar{y}$ is the mean measurement response

In case the adequateness of the linear model cannot be proven, the non-linear second order calibration function model should be applied. In this case the graphical representation of the data is a parabola.

In the non-linear model the main parameters (a, b, c) are defined by the equation:

$y=a+bx+cx^2$

Where:

x is the concentration

y is the area value

a is a constant term, it controls the height of the parabola, more specifically, it is the point were the parabola crosses the *y*-axis. It is a constant term.

b is the linear coefficient

c is the quadratic coefficient, the declivity of the parabola as it crosses the y-axis

The coefficients *b* and *c* together control the axis of symmetry of the parabola (also the *x*-coordinate of the vertex).

The data from the 6 calibration curves derived from the 2 repetitions on the 3 different days in the reproducibility assessment can be used for the assessment of the linearity model (see Figure 1 – flow chart of full validation scheme and worked example from paragraph 5.2.7.1.3)

5.2.3.3 Assessment of the linearity model by fixed predefined acceptability criteria

5.2.3.3.1 Maximum allowable standard deviation of the slope

The maximum relative standard deviation of the slope, calculated according to the following formula:

s_b (%) = (s_b/b) x 100, where:

s_b = standard deviation of the slope from several replicates of the calibration curves in repeatability conditions and

b = slope

The standard deviation of the slope should not exceed 5% for classical chromatography techniques (GC-FID, HPLC-UV, DAD, FLD, etc.) and 8% for more specialised techniques (MS detection).

5.2.3.3.2 Maximum allowable residue for different calibration levels

The maximum acceptable residues for each calibration level is method dependent, so it should be specified during the validation study and should be set up as a criteria for future assessment of the calibration curves produced in routine analysis[1]. The maximum acceptable residue for the first level of calibration could be much higher for the other levels.

[1] As an example a typical maximum acceptable residue for LC/MS analysis could be

- the residue calculated for the lowest level of calibration curve (at LOQ level) < 20%;
- the residues calculated for all other levels of calibration < 15%, expressed with the following Figure 5

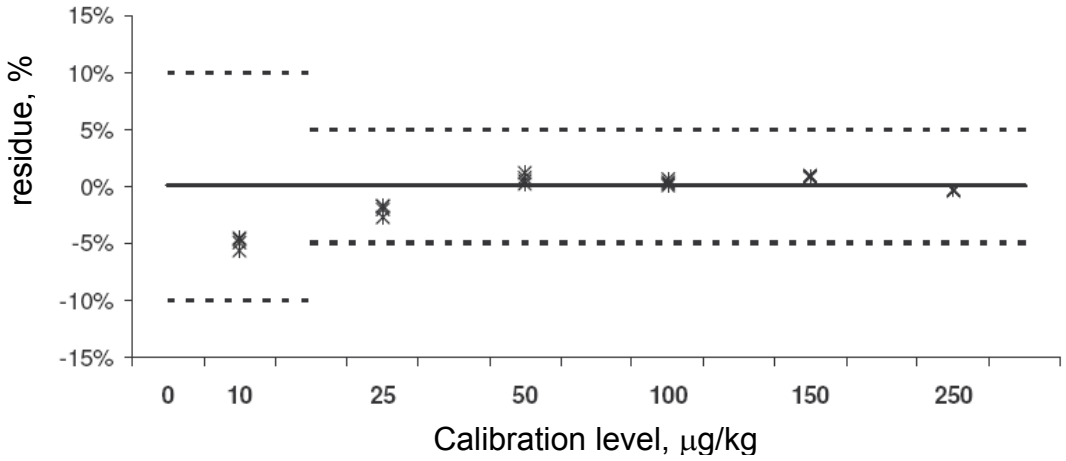


Figure 5: Plot of residues versus concentration

5.2.3.4 Assessment of the adequateness of the linearity model by statistical tests

To verify the linearity of the model a number of different statistical tests can be used. The most commonly used are described in the following paragraphs.

5.2.3.4.1 ANOVA for lack of fit (ISO approach, [11])

The test for adequateness of the linearity model allows the validity of the regression model and the chosen working range to be verified. The ANOVA lack of fit model is based on the comparison of the tabulated F of Fisher values with the observed F of Fischer calculated on the basis on the experimental results, and on the sums of squares.

To verify the linearity, the sum of the squares of the difference between any measured value y_{ij} and the general mean $\bar{\bar{y}}$ is calculated from the sum of squares of the linear regression and the sum of the squares due to the error of the model (and therefore a non linearity) , as follows:

$$\sum_i \sum_j (y_{ij} - \bar{\bar{y}})^2 = \sum_i \sum_j (\hat{y}_{ij} - \bar{\bar{y}})^2 + \sum_i \sum_j (\bar{y}_i - \hat{y}_{ij})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 \qquad 1 \le i \le p$$

Where:

$\hat{y}_{ij}$ is the response predicted by the model;

$\bar{y}_i$ is the average of the replicates

$SSD_{total} = SSD_{cal.function} + SSD_{residual} = SSD_{cal.function} + SSD_{lack\ of\ fit} + SSD_{pure\ error}$

where:

$SSD_{cal.function} = SSD_{total} - SSD_{residual}$

$SSD_{lack\ of\ fit} = SSD_{residual} - SSD_{pure\ error}$

$SSD_{pure\ error} = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$

$SSD_{total} = \sum_i \sum_j (y_{ij} - \bar{\bar{y}})^2$

$SSD_{residual} = \sum_i \sum_j (y_{ij} - \hat{y}_i)^2 = \sum_i \sum_j (\varepsilon_{ij})^2$

$SSD_{\text{lack of fit}}$ represents the sum of the squares due to the error of the model (non-linearity)

SSD_{residual} represents the sum of squares due to the distance between the average calculated in the sample considered representative of the population (working range, calibration curve range) and the real average value that should be calculated in the total population (considering an infinite number of calibration points)

The observed F values relative to the calibration function ($F_{\text{cal.function}}$) and to the lack of fit ($F_{\text{lack of fit}}$) are calculated on the basis of the ratio between the variance due to the linearity and that due to non linearity, and the variance due to the residue, as follows:

$$F_{calfunction} = \frac{s_{calfunction_l}^{2}(y)}{s_{pureerror}^{2}(y)} \qquad \text{and} \qquad F_{lackoffit} = \frac{s_{lackoffit}^{2}(y)}{s_{pureerror}^{2}(y)}$$

Where:

$$s_{cal.function}^{2}(y) = \frac{SSD_{cal.function}(y)}{1}$$

$$s_{lack.of.fit}^{2}(y) = \frac{SSD_{lackoffit}(y)}{p-2}$$

$$s_{pure.error}^{2}(y) = \frac{SSD_{pure.error}(y)}{p(n-1)}$$

Where:

n is the number of replicates, which means 1 for each day of reproducibility study, for a total of 3. Each one of these 3 values corresponds to a mean of 2 values.

p is the number of calibration levels (at least 5 plus 1 blank)

(p-2) represents the number of degrees of freedom relative to the error of the model (non-linearity)

p(n-1) represents number of degrees of freedom relative the residual.

s^2 is the variance

For the calculation of degrees of freedom associated with any of the expressions, the following formula can be used:

Total = 1+ Error of the model (non-linearity) + Residual

$(np - 1) = 1 + (p-2) + p(n-1)$

In Table 5, the representative values that have to be calculated for the evaluation of the linearity are reported.

Variation source	Sum of squares	Degree of freedom	Variance	F of Fisher (variance ratio)	F_1 Fisher 5%	F_2 Fisher 1%
Calibration function	SSD $_{cal.function}$	1	$\frac{SSD_{cal.function}(y)}{1}$	$F_{cal.function} = \frac{s^2_{cal.function}(y)}{s^2_{pure.error}(y)}$	F_{tab} (1, (p*(n-1), 95%)	F_{tab} (1, , (p*(n-1), 99%)
Lack of fit	SSD $_{lack\ of\ fit}$	*(p-2)*	$\frac{SSD_{lack.of.fit}(y)}{p-2}$	$F_{lack.of.fit} = \frac{s^2_{lack.of.fit}(y)}{s^2_{pure.error}(y)}$	F_{tab} (p-2, , (p*(n-1), 95%)	F_{tab} (p-2, , (p*(n-1), 99%)
Pure error	SSD $_{pure\ error}$	*(p*(n-1))*	$\frac{SSD_{pure.error}(y)}{p(n-1)}$			
Total	SSD $_{total}$	*(p*n-1)*				

Table 5: ANOVA table to compare lack of fit and pure error under the assumption of proportional residual standard deviation

The data from the calibration curves coming from the 2 repetitions in the 3 different days (reproducibility assessment) can be used for the assessment of linearity, following the ANOVA lack to fit approach.

5.2.3.4.2 Plot of the residuals

The plot of the residuals ε versus the corresponding concentration (x) values or the fitted value $\hat{y}_i$ is a powerful tool to verify the two assumptions of linearity and of constant residual standard deviation.

If these two assumptions are confirmed, then the figure should display a plot of randomly distributed points centred on zero (as shown in Figure 6).

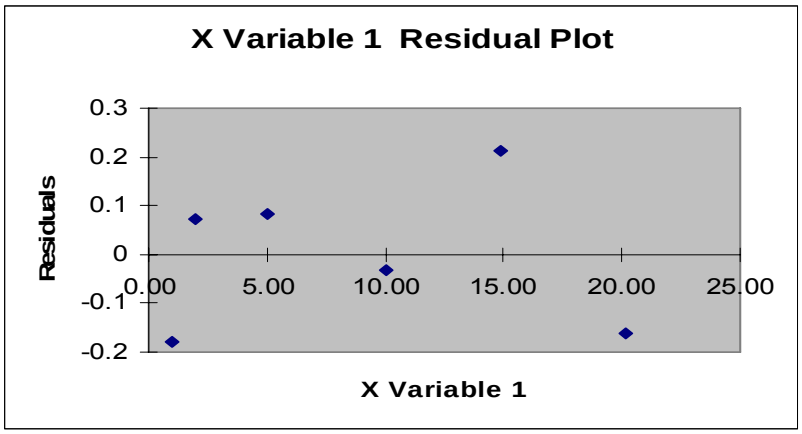


Figure 6: Residual plot

If the model is not linear, a systematic pattern between the residuals and the concentration values can be seen (as shown in Figure 7).

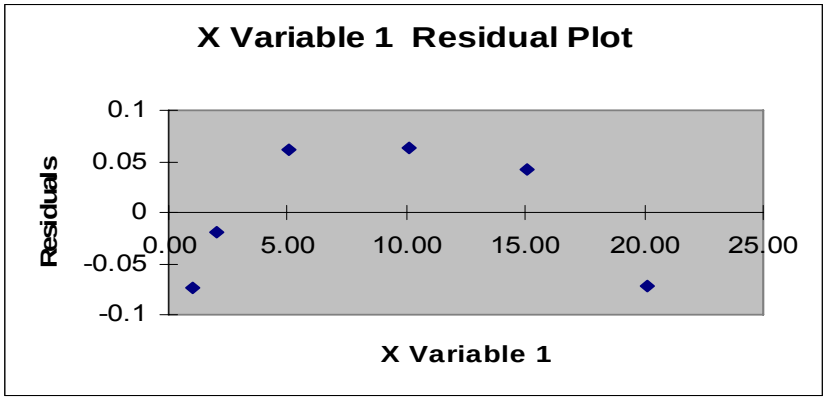


Figure 7: Residual plot

Departure from the assumption of constant residual standard deviation is indicated by dispersion in the data that increases or decreases with the fitted value.

If the assumption of constant residual standard deviation does not hold, then the data collected during the calibration experiment must be re-analysed.

5.2.3.4.3 Acceptability criteria for ANOVA for lack of fit approach

5.2.3.4.3.1 Validity of the regression model

If the F of the Fisher value, relative to the linear regression (F_i), calculated on the basis of the experimental results, is higher then the F of the Fisher value tabulated (8.53), calculated on the basis of 6 degrees of freedom at the numerator and 16 degrees of freedom at the denominator, as reported in table 1a in Annex 1, the regression model could be considered as acceptable at the risk level of 1%. If F_i is between 8.53 and 4.49 (see Tables 1a and 1b in Annex 1) the model can be considered as acceptable at the risk level of 5%. If F_i is lower then 4.49 (see Table 1b in Annex 1) the model cannot be used to prove linearity.

5.2.3.4.3.2 Validity of the chosen calibration range

If the non-linearity F value (F_{ni}) is lower than the F of the tabulated Fisher value (2.74), calculated on the basis of 6 degrees of freedom at the numerator and 16 degrees of freedom at the denominator, as reported in Table 1b in Annex 1, the chosen working range is validated, within the possible error at level of 5%. If the value is between 2.74 and 4.20 (see Tables 1a and 1b in Annex 1), the chosen working range is validated, with the possible error at level of 1%.

5.2.3.4.4 Significance of the quadratic coefficient

The significance of the quadratic coefficient can be checked.

If the confidence interval (CI, the interval into which the value of c will fall with a probability P, generally 95%) of the quadratic coefficient c (from the equation: $y = a + b*x + c*x^2$, interpolating the calibration curve) includes zero, the quadratic term becomes negligible and the equation is reduced to a linear function $y = a + b*x$.

The confidence interval can be calculated with the following formula:

$CI_c = t_{(P,n-3)} * s_c$

Where:

CI_c confidence interval of the quadratic coefficient

$t_{(P,f)}$ Student's t value for level of statistical confidence P and degrees of freedom f $t_{(P,n-3)}$

s_c is the standard deviation of the quadratic coefficient c, based on the different calibration curves calculated.

If $c \pm CI_c$ includes 0: the quadratic coefficient is not significant, i.e. there is no significantly better fit using the quadratic regression, so the linear regression (of the type: $y = a + b*x$) can be accepted.

The data from the calibration curves coming from the 2 repetitions in the 3 different days (reproducibility assessment) can be used for the assessment of linearity, following the present approach.

In addition to the linear regression model the data also have to be interpolated with a quadratic regression, $y = a + b*x + c*x^2$. The standard deviation (s_c) and confidence interval (CI) of the 6 values of the quadratic coefficient c obtained have to be calculated and the statistical test described above has to be applied.

5.2.3.4.4.1 Acceptability criteria for non significance of quadratic coefficient

If $c \pm CI_c$ includes 0, then the quadratic coefficient is not significant and the term cx^2 can be omitted. This means that the quadratic regression does not represent a better fit for the calibration curve than a linear regression.

In this case the linear regression (of the type: $y = a + b*x$) can be accepted.

5.2.3.4.5 Mandel test

The test according to Mandel [10] is based on the comparison of a linear regression and a quadratic regression to interpolate the data coming from the analysis of the calibration curve.

In this case, a test is performed to determine whether the quadratic function represents a significantly better model than the linear function, by comparing the residual variances of both regressions.

For this purpose a value (F_{calc}) the effects of the residual variance of the linear regression (s_{y1}^2) and the residual variance of the quadratic regression (s_{y2}^2) are combined and compared with the tabulated F of Fisher (for a level of statistical confidence P (generally 95%) and degrees of freedom 1 and n-3).

Thus:

$F_{calc} = [(n-2)* s_{y1}^2 – (n-3)*s_{y2}^2] / s_{y2}^2$

Where:

n is the number of the calibration curves used for this linearity verification

s_{y1}^2 is the residual variance of the linear regression

s_{y2}^2 is the residual variance of the quadratic regression

The data from the calibration curves coming from the 2 repetitions in the 3 different days (reproducibility assessment) can be used for the assessment of linearity, following this approach.

Calculations for residual variance are performed as described in 5.2.3.4.1.

5.2.3.4.5.1 Acceptability criteria for Mandel test

If F_{calc}< F, then the quadratic regression is not significantly better than the linear regression and thus the linear regression can be accepted.

5.2.3.5 Choice of calibration with or without weighing

The homogeneity of variances over the whole range (homoscedasticity, the variance of the lowest level is comparable with the variance of the highest level) is a prerequisite for a non-weighted linear regression [11]. Thus, if the model has demonstrated that the model is linear, but the responses are not normally distributed or the variances within the calibration range are not homogeneous and the working range is over more than two orders of magnitude, there is the need to decide whether to apply the weighting factor to the linear regression.

The homogeneity can be verified by calculating the standard deviation of repeated measurements (n = 6 to 10) at the lowest and at the highest levels of the required linear range. There are two possibilities:

The standard deviation is not significantly different: weighing is not necessary.

The standard deviation is significantly different: weighting is necessary and the weighting factor should be selected.

To evaluate this, the statistical F-Test is applied and the F value relative to the standard deviations of the lowest and the highest calibration levels have to be calculated and compared with the tabulated F of Fisher for probability P (generally 95%) and degrees of freedom (n-1, n-1). To calculate F_{calc} the following formula should be used:

$$F_{calc} = s^2_{low}/s^2_{high}$$

s^2 are the variances of the replicates at the low and the high calibration levels

If F_{calc} < F: The standard deviation is not significantly different: weighting is not necessary.

If F_{calc} > F: The standard deviation is significantly different: weighting is necessary and the weighing factor should be selected.

If the standard deviation is significantly different, the interval of the calibration curve could be reduced, or split into narrower calibration ranges (for example for calibration range of 0 and 1000 could be split into three different ranges of 0-10, 10-100 and

100-1000). Otherwise if weighting is shown to be necessary, as the interval cannot be reduced or split, the concentration values should be calculated by using different weighing factors.

The most commonly used factors are:

1/x;

$1/x^2$;

1/y or

$1/y^2$

Where:

x is the response of the instrument

y is the concentration calculated based on the calibration curve

The results should be calculated as % recovery (calculated value / real value *100%). As well as the % recovery the relative error (the calculated difference between the % recovery and 100%) should be determined for all the levels of the calibration curve and all weighting factors. Then the sum of all of the absolute relative errors at all levels of the calibration curve for each weighing factor should be calculated.

The calibration curve calculated with the weighting factor that provides the smallest summarised relative error should be used.

The data from the calibration curves coming from the 2 repetitions in the 3 different days (reproducibility assessment) can be used for the decision if weighting has to be applied to the calibration curves.

5.2.4 Working concentration range

The working concentration range is the range in which the method is validated and which gives an acceptable trueness and precision. It should be distinguished from the calibration range in which the regression model for calibration (most frequently linear one) is established and verified.

The lowest limit of the working range is the lowest limit of the calibration range. However the upper limit of the working range could be not only the highest point in the calibration curve, for which the regression model is validated, but a higher concentration at which acceptable trueness can be proven (e.g. by comparing with t-test the results of a diluted upper limit concentration with a concentration inside the calibration range).

5.2.5 Limit of detection (LOD) / method detection limit (MDL)

The limit of detection (LOD) is the smallest measured concentration of an analyte, which allows the presence of the analytes to be detected in the test sample with

acceptable certainty.

When the level of interest or the legislative limit (LL) is much higher than the detection limit, the determination of the detection limit is of no value. The limit of detection does not need to be calculated (and thus reported in the validation report) when:

- the lowest concentration point of the calibration curve should be 10 times lower than the LL, if possible or at least 5 times lower than the LL, corresponding to LOQ level;
- the signal to noise ratio for concentration level corresponding to the LL is higher than 30:1.

The calculation of detection limit is very important when the validation is performed for substances that have to be “non detectable” with a defined detection limit, as the capability of the method to reach the target detection limit should be proven.

In this case a blank sample has to be spiked with the concentration considered as limit of detection and the verification that it can be clearly distinguished from the blank has to be performed. This means that 3 repeated analyses have to be performed and if the mean response of the spiked samples (at the limit of detection level) is greater than 3 times the maximum blank value the limit of detection is verified.

It is important to distinguish between the instrument detection limit (IDL) and the method detection limit (MDL). The IDL is an instrument parameter and can be obtained from measurements of pure analyte, in contrast to the MDL (LOD) which is based on measurements of blank real sample or a low-level spiked sample that has been taken through all the steps of the method.

The limit of detection can be calculated in several ways and the most used are reported in the next paragraphs.

5.2.5.1 Calculation from standard deviation of the blank [12]

This approach (IUPAC approach) is usually performed with instrumental methods.

Blank samples have to be analysed according to the method, and their standard deviation calculated.

The number of analyses to be performed must be in any case equal to or higher than 6 ($n \geq 6$).

The detection limit is defined as:

$$LOD = \bar{x}_{BL} + 3.s_{Bl}$$

Where:

$\bar{x}_{BL}$ is the mean concentration calculated from the area of a noise peak for at least 6 analyses of a blank sample. In case the noise peak in the blank sample is not measurable, the peak relative to the lowest concentration of the calibration curve can be used, provided its value is comparable to LOD. In this case the LOD will be simply 3 times the standard deviation of the lowest concentration level, as $\bar{x}_{BL}$ is considered as zero.

s_{BL} is the standard deviation of the analysis.

The blank samples used to calculate the LOD are the same as those prepared in the reproducibility assessment. No additional analyses have to be performed.

The LOD should be confirmed by spiking a blank sample at this level to see if it is really visible with S/N of at least 3.

5.2.5.2 Calculation form the signal to noise ratio

This approach is followed for analytical methods producing a baseline response. For this reason it is widely applied for chromatographic methods. The signal to noise ratio (S/N) is determined by comparison between the signals obtained from samples containing known low concentrations of the analyte and the signal obtained from a blank sample. The lowest analyte concentration level that can be determined with acceptable reliability is calculated. The limit of detection is defined as the concentration of the analyte at a signal/noise ration S/N=3 after blank correction

5.2.5.3 Calculation from standard deviation of the intercept [12]

If calibrations are used to quantify the analyte, the intercept can be regarded as an extrapolation of a blank determination.

The detection limit can then be defined as 3 times the standard deviation of the intercept (the b value in the equation y=ax+b), according to the formula:

$$LOD = 3.s_b$$

Where:

s_b is the standard deviation of the b value of the intercept

The LOD values obtained with the formula above are then converted to concentration using the calibration curve.

The calibration curves used to calculate the LOD in this way are the same as those prepared in the reproducibility assessment. No additional analyses have to be performed.

5.2.5.4 Calculation according to the German standard DIN 32645 [35]

This approach is also based on the regression analyses of the calibration curves, so only data from the calibration curves are used.

These calculations are performed by the instrument software (Chemstation etc.), LIMS or through others specific software (DINTEST, Valistat, Effichem, MVA3, etc.).

In DIN 32645 [35] it is remarked, that the calculation of LOD/LOQ with the calibration function should based on calibration points close to the LOD/LOQ. If it is done from 10 times the LOD or using the calibration from working range, it results in false findings.

5.2.6 Limit of quantification

The limit of quantification represents the lowest concentration of analyte that can be determined with an acceptable level of uncertainty e.g. the lowest reportable result. It should be established by using an appropriate reference material or sample. It should be not determined by extrapolation, although by conventions it is usually taken as a fixed multiple (2-5) of the detection limit. This approach gives an approximate value; its variation with the type of sample should be taken in consideration, so a verification of the calculated LOQ is required.

For the present method validation scheme the limit of quantification x_{LOQ} should be calculated as:

$x_{LOQ} = 2\ x_{LOD}$

5.2.6.1 Verification of the limit of quantification

Verification of the limit of quantification has to be based on the level of standard deviation acceptable for the method under validation.

A blank matrix sample has to be spiked with the concentration considered as limit of quantification and analysed at least 5 times under reproducibility conditions. The mean value ($\overline{x}_{LOQ}$) and the standard deviations (s_{LOQ}) of these repeated analyses have to be calculated.

These values establish whether or not the precision of the calculated limit of quantification is lower than the fixed acceptable limit (for example 30% s_{LOQ} as from convention in the field of FCM).

5.2.7 Accuracy

Accuracy expresses the quality of the method. Method validation seeks to quantify the likely accuracy of results by assessing both systematic and random effects on the results. Accuracy is, therefore, normally studied as two components - trueness and precision.

Precision is a measure of how close results are to one another and is usually expressed by measures such as standard deviation (SD) or as relative standard deviation (RSD) sometimes called the coefficient of variation (CV), which describe the spread of results. Precision is generally dependent on analyte concentration, and so should be determined at a number of concentrations and if relevant, the relationship between precision and analyte concentration should be established. In that case relative standard deviation is more useful. Precision is normally determined for specific circumstances, which in practice can be very varied. The two common precision measures are repeatability and reproducibility.

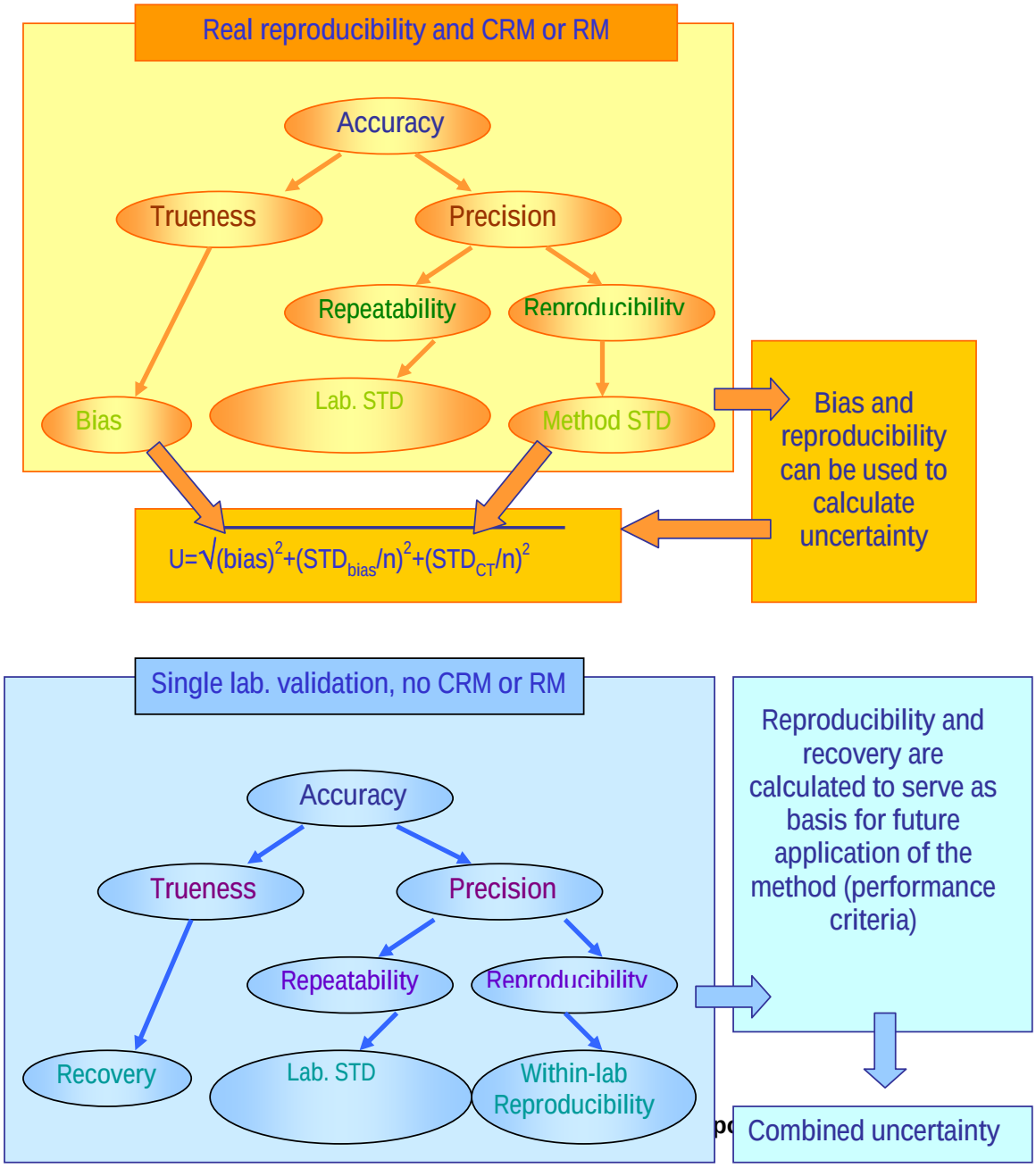
Trueness of a method is an expression of how close the mean of a set of results (produced by the method) is to the true value. Trueness is usually expressed in terms of bias, the difference between the true value (calculated on the basis of a certified reference material or a reference material) and the mean of the results.

When certified reference materials or reference materials are not available, trueness can be expressed in terms of recovery, after spiking.

Uncertainty of measurement is another common expression of accuracy. It is calculated combining bias and its relative standard deviation (calculated for the assessment of trueness) and the standard deviation of the reproducibility assessment.

When certified reference materials or reference materials are not available no real bias can be assessed. Further for single laboratory validation only a within laboratory reproducibility can be determined. Thus, in these cases uncertainty cannot be calculated based on bias and reproducibility, and it has to be estimated through the combination of all possible sources of uncertainty associated with the method (combined uncertainty).

Figure 8 represents a graphical representation of the accuracy components and on how to assess them.



5.2.7.1 Precision

5.2.7.1.1 Definitions

Repeatability means precision under repeatability conditions (r), which means the closeness of agreement between mutually independent test results obtained under repeatability conditions, i.e. using the same method on identical test material in the same laboratory by the same operator using the same equipment within short intervals of time, often called for that short-term repeatability.

The repeatability limit value obtained in the validation study is intended as a reference for the general use of the method: each laboratory using the method during their daily work can use the established repeatability limit to verify the appropriateness of their analyses.

Reproducibility means precision under reproducibility conditions (R), which means the distribution of measurement results obtained under reproducibility conditions. It may be measured by means of collaborative studies (identical test materials, identical methods, different persons, different instruments and different laboratories). It is expressed as standard deviation and it represents the overall precision of the method.

For a single laboratory validation (SLV) scheme, real reproducibility cannot be assessed, but only within-laboratory reproducibility (intermediate precision) is determined. It is expressed as within-laboratory standard deviation or CV (%) of different batches of samples analysed in the single laboratory by different operators, using different instruments over long period of time, varying as much as possible the factors influencing the analytical result and representing the real fluctuation in performing a method in the laboratory. It represents the precision of the method but only for the laboratory that performed the validation.

5.2.7.1.2 Short-term repeatability

The determination of repeatability can be basically divided into three parts:

- Collection of experimental data;
- Evaluation of experimental data;
- Statistical processing of experimental data.

5.2.7.1.2.1 Experimental Design / Collection of experimental data

The number of experimental data and thus of experimental trials to be performed to obtain such data is dependent on:

- The choice of the number of analyte concentration levels, considering operating range of the method;
- The choice of the number (n) of replicate tests to be performed for each analyte concentration level, considering the standard deviation reliability according to UNICHIM Manual No. 179/1, section 6 [6];

For each concentration level, the stipulated number of tests on the same sample is

performed.

For spiked samples, at least 6 replicate analyses per sample spiked at a minimum of 3 concentration levels within the working range (e.g. 0.2 LL, 1*LL and 2*LL) have to be performed under the same conditions, by the same person and with the same equipment, representing minimum variability (repeatability conditions).

If a CRM or RM is available at least 6 replicate analyses for each available concentration level have to be performed under the same conditions as for spiked samples. The number of the concentration levels in this case depends on the number of available CRM or RM concentrations (normally only one). Thus, the method can be validated only in a range very close to the available concentrations. In case a broader range of concentrations is necessary, additional studies on spiked samples have to be performed.

Worked example for spiked samples

Samples: A total of 24 analyses have to be performed (6 times 3 concentrations + blank each one) by one operator.
=> 24 analyses in day 1

Calibration curve 1, 2: A total of 12 analyses had to be performed (2 times 5 concentrations+ blank) by one operator.
=> 12 analyses in day 1

=> 36 analyses in day 1 by operator A

5.2.7.1.2.2 Statistical evaluation of experimental data

For each series of n data obtained for each analyte concentration level the following steps should be carried out:

- Verification of the normal data distribution with the Shapiro-Wilk Normality test;
- Normal distribution has to be verified according Shapiro-Wilk test, as indicated in [19] and in UNICHIM Manual No. 179/1, section 7.1 [6] separately for all the 9 replicate batches. The normality hypothesis can be accepted with Kp < [K p=1-α] for p=1-α=0.95
- If the distribution is proven to be normal, the evaluation has to proceed with the next tests. If anomalous values were found, first they have to be eliminated and then the normality test repeated. In the absence of anomalous data or in the event that the distribution remained anomalous even after they have been eliminated, the whole process must be reviewed, in order to identify the cause(s) of the anomaly.
- A verification of the presence of anomalous data should then be performed, by applying the Grubbs or Dixon's tests to each series of data, as indicated in UNICHIM Manual No. 179/1, sections 8.1 and 8.2 [6] and in [4], for a significance level α equal to 0.05. If anomalous data are present, they have to be eliminated, endeavouring to trace the cause or causes that produced them. Once the anomalous data have been rejected, the Shapiro-Wilk Normality test has to be repeated.

After the verification of the normal distribution of the sets of data and the rejection of the anomalous data (outliers), the data should be processed according to the procedure described in the following paragraph.

5.2.7.1.2.3 Processing of experimental data

After the evaluation of the consistency of the data sets, for each analysed concentration level (series of n data obtained), the following calculations had to be performed:

- Mean value
- Estimation of the standard deviation (sr), and/or
- Coefficient of variation (CVr) or repeatability relative standard deviation RSDr.

From the repeatability relative standard deviation it is useful to calculate the "repeatability limit" r, which is, according to ISO Standard 78-2 [20], the value less than or equal to the absolute difference of two test results which can be expected with a probability of 95%, under repeatability conditions. It enables the analyst to decide whether the difference between duplicate analyses of a sample (Δx), determined under repeatability conditions, is significant. Thus, if the difference between 2 results Δx exceeds r, the results should be considered as suspect.

Once the r value is fixed based on the results of the validation study, then for all the subsequent applications of the method it should always to be higher than the difference between duplicate analyses. This allows the analysts that will use the validated method to verify the consistency of their results.

For this reason, the value of r should be calculated as follows:

$$r = t * \sqrt{2} * s_r$$

where t = 2.45 for 6 replicates for *s* calculation and n=2 (for difference between two results).

In Table 2, in Annex 1 the student t-values with the relative effective degrees of freedom are reported.

5.2.7.1.3 Within-laboratory reproducibility (intermediate precision)

5.2.7.1.3.1 Experimental Design/ Collection of experimental Data

In order to assess within-laboratory reproducibility (intermediate precision) in SLV study, the largest possible within-laboratory variability should be introduced, performing non simultaneous replicates conducted in the same laboratory on identical test samples on different days, by different analysts, with different instruments, using different calibration curves, and with different sources of reagents, solvents and columns.

To assess within-laboratory reproducibility, the whole analytical method has to be repeated on 3 sets of samples in the low, middle and high concentration range at least in 3 different days encompassing all possible within-laboratory variability conditions: repeated by different operators (if they are not available at least two) in different days and in different environmental conditions (if they constitute a critical parameter) with different instruments (if possible).

If CRM, RM or internal quality control samples (QC samples) are available at least 6

repeated analyses for each available concentration level have to be performed under reproducibility conditions. The number of the concentration levels in this case depends on the number of available CRM, RM or QC samples concentrations (normally only one). Thus, the method can be validated only in a range very close to the available concentrations. In case a broader range of concentrations is necessary, additional studies on spiked samples have to be performed.

For spiked samples, the 3 concentration levels should be: lowest calibration point as defined according to paragraph 5.2.3.1, LL, and 2-5*LL. At least 6 replicate analyses should be performed for each of the 3 concentration levels have to be performed under the reproducibility conditions.

Worked example on spiked samples

Samples: In addition to the samples analysed in the assessment of the mehod repeatability another two sets of analyses have to be carried out (not in sequence but on another 2 different days). The same samples described in the repeatability assessment should be prepared on the two additional days.

48 analyses in day 2 and 3 (24 in day 2 by operator A and 24 in day 3 by operator B)

Calibration curve day 2 and day 3: Another two sets of calibration samples should be prepared (not in sequence but on another 2 different days), for a total of 24 new analyses (5 concentrations + blank for 2 batches of samples -1 set was already analysed to assess repeatability) per day, by two different operators.

24 analyses in day 2 and 3 (12 in day 2 by operator A and 12 in day 3 by operator B)

72 analyses in day 2 and 3 (36 in day 2 by operator A, and 36 in day 3 by operator B)

5.2.7.1.3.2 Statistical evaluation of experimental data

Before processing the raw data, they should be statistically evaluated for:

- Normality (Shapiro-Wilk Normality test) according to the procedure described at 5.2.7.1.2.2.
- Presence of anomalous data by the Dixon and/or Grubbs tests, according to the procedure described in 5.2.7.1.2.2.
- Homogeneity of variance - Homogeneity of variance between different batches/days for the same level is verified by performing the Cochran and minimum variance tests, for the case of data series with the same number of samples and a significance level α equal to 0.05, as indicated in the UNICHIM Manual No. 179/2, sections 8.1 and 8.2 [6] and [4]. In case of different number of results per series, the Bartlett and Hartley test could be applied.

Cochran's test is based on homogeneity of variance, under repeatability conditions. It tests only the highest value in a set of standard deviations or ranges and is therefore a one-sided outlier test. If the test statistic on a certain item lies between its 5% and 1% critical values, then the entry is called a statistical straggler; if the test statistic is greater than its 1% critical value then the item is called a statistical outlier. The critical values for Cochran's test are reported in Annex A of ISO 5725-2:1994 [4].

Eliminate data series for which the variances are significantly different from others, endeavouring, if possible, to trace the causes of the anomaly.

- Homogeneity of mean values - Verification of the homogeneity of the mean values by application of variance analysis (ANOVA = analysis of variance).

- The compatibility of the various means via the single-factor variance analysis (ANOVA), has to be verified for a significance level α equal to 0.05, as indicated in [18] and UNICHIM Manual No. 179/2, sections 9.1 and 9.2 [6].

If there are incompatibilities between the mean values, a Scheffè test has to be performed for a significance level α equal to 0.05, as indicated in [19] and the UNICHIM Manual No. 179/2, section 10 [6] and to eliminate data series for which the mean values are significantly different from the others.

ANOVA has to be applied with the hypothesis (H_0) that the reproducibility conditions have no influence on the final results.

In ANOVA F is the statistic value that represents the ratio of two variance estimates:

F-RATIO = Between-group Variance/Within-group Variance

If H_0 is true: F = (0+ Error)/ Error ≅ 1

If H_0 is false: F = Treatment Effect + Error)/Error > 1

If H_0 is the true F value calculated on the basis of the experimental results then it has to be lower than the F critical value, calculated by ANOVA.

If F calculated, is higher than F critical, then the reproducibility conditions have an influence on the performance of the method and thus the method cannot be considered reproducible.

If the null hypothesis (H_0) is rejected by ANOVA this implies that at least one of the means isn't the same as the other means, so the reason should be found. For this the Scheffè test is used, testing if pairs of means are different one from the other.

5.2.7.1.3.3 Processing of experimental data

After having assessed that the experimental data are homogeneous and that the reproducibility conditions have no influence on the final results, data have to be statistically processed separately for each analyte concentration level, to obtain:

- the general mean value $\bar{\bar{x}}$ is calculates as a mean value from all the $\bar{x}_i$ statistically evaluated experimental data;
- the estimation of the within-laboratory standard deviation (s_{WR}):

$$s_{WR} = \sqrt{s_r^2 + s_{bb}^2}$$

where:

s_r is within batch standard deviation representing repeatability s :

$$s_r = \sqrt{\frac{\sum_{i=1}^{p} s_{r,i}^2}{p}}$$

where p is the number of batches,

s_{bb} is the between batch standard deviation

$$s_{bb} = \sqrt{s(\bar{x}_i)^2 - \frac{s_r^2}{n}}$$

where n is the number of replicates within a batch.

In case when $s(\bar{x}_i)^2 - \frac{s_r^2}{n} < 0$, then

$s_{WR} = s_r$

- The coefficient of variation CV_{WR}(%) of the standard deviation as reproducibility relative standard deviation RS_{WR} (%);
- The within laboratory limits of reproducibility, WR.

The within laboratory reproducibility limit was calculated for each analyte level, using the following equation:

$$WR_{LAB} = t * \sqrt{2} * s_{WR}$$

where

t is the Student's t-value and its values are: t=2.11 for 18 replicate for S calculation and n=2 (for difference between two results).

In case of problems with the analyses resulting in only 17 replicates (i.e. one of the replicates was excluded as an outlier) then t=2.12 should be used. All t-values should correspond to a 97.5% probability value.

The "within laboratory reproducibility limit" WR enables the analyst to decide whether the difference between two analyses of sample (Δx), determined under reproducibility conditions within the same laboratory, is significant.

Once the WR value is fixed based on the results of the validation study, then for all the subsequent applications of the method it should always be higher than the difference between duplicate analyses. This will allow the analysts that will use the validated method to verify the consistency of their results.

5.2.7.1.4 Acceptability criteria for precision

The within-laboratory standard deviation s_{WR} for the repeated analysis of a reference or fortified material, under reproducibility conditions, should not exceed the level calculated by the Horwitz Equation (Table 6).

Analyte %	Analyte ratio	Unit	RSD (%)predicted
0.01	10-4	100 ppm	8.0
0.001	10-5	10 ppm	11.3
0.0001	10-6	1 ppm	16.0
0.00001	10-7	100 ppb	22.6

Table 6: Predicted value for within-laboratory laboratory RSD depending on concentration

The equation is:

$RSD\ (\%)_{predicted} = 2^{(1-0.5\log C)} = 2*C^{(-0.5\log 2)} = 2*C^{-0.15}$

Where:

C is the mass fraction expressed as a power (exponent) of 10 (e.g. 1 mg/g = 10-3).

A more contemporary model based on results from PT schemes has shown that the relationship is best represented if three equations are used to cover from high to low concentrations.

For concentration lower than 1.2×10^{-7} and higher then 0.138 Thompson [17] introduced a correction in the Horwitz formula:

$s\ (\%)_{predicted} = 0.22*C,$ if $C < 1.2 \times 10^{-7}$

$s\ (\%)_{predicted} = 0.01*C^{0.5}$ if $C > 0.138$

For analyses carried out under short term repeatability conditions, the repeatability standard deviation s_r would typically be between one half and two thirds of the above values.

5.2.7.2 Trueness

5.2.7.2.1 Definitions

Trueness is the degree of agreement between a sample's true content of a specific analyte and the result of the analysis.

The sample's true content of an analyte is always unknown. In order to evaluate the trueness of a method, it is therefore necessary to depend on accepted “reference values”

Therefore, the trueness is the closeness of agreement between a test result and the accepted reference value.

Trueness is stated quantitatively in terms of “bias”, with smaller bias indicating greater trueness.

Bias (%) = (mean concentration - reference value) x 100/reference value.

Bias can arise at different levels in an analytical system, for example, run bias, laboratory bias, and method bias.

Bias is typically determined by comparing the response of the method to a reference value

Sources of reference values for trueness experiments are: CRMs, methods for which organised collaborative studies and proficiency testing results are available, reference materials (RM), reference methods, and reference laboratories.

Reference value	Uncertainty of the reference value
X_{CRM} (CRM)	$U_{REF} = U_{CRM}/2$ if in the certificate the uncertainty of the reference value is expanded uncertainty with coverage factor of 2 if not $U_{REF} = U_{CRM}/ SQRT(3)$
X_{RM} (Reference material)	$U_{REF} = U_{RM}/2$ if in the certificate the uncertainty of the reference value is expanded uncertainty with coverage factor of 2 if not $U_{REF} = U_{RM}/ SQRT(3)$
$X_{Interlab\ comparison}$	$U_{REF} = U_{IL}$ provided by the organisers of an ILC or $U_{REF} = s_{interlab\ reproducibility}/SQRT(p)$; p=number of laboratories
X_{RM} (Reference method)	$U_{REF} = s_{RM}$ from the n results obtained with the reference method/ SQRT(n)
X_{RL} (Reference Laboratory)	$U_{REF} = S_{RL}$ from the n results obtained by the reference laboratory/ SQRT(n)

Table 7: Different calculation of the uncertainty based on different reference value type

When CRMs, reference methods or collaborative study or proficiency test results are not available the results of the spiking/recovery experiments can be used (see section 5.2.7.2.2.6).

5.2.7.2.2 Estimation of trueness

5.2.7.2.2.1 Certified Reference Materials (CRMs)

They should be natural matrices, closely similar to the sample of interest for which the method is under validation. CRMs are traceable to international standards, with a known uncertainty and therefore can be used to assess bias, assuming that there is no matrix mismatch. CRMs should always be used in validation of trueness where it is possible to do so. At the present time, in the field of FCM only a small number of CRM are available.

For trueness estimation the following steps have to be performed:

- Analyse six replicates of the CRM with different concentration levels of the analyte of interest (if available) in accordance with the method instructions
- Determine the concentration of the analyte present in each sample of the replicate samples
- Perform the statistical evaluation of the data through Shapiro-Wilk Normality test and Grubbs outliers test (as in 5.2.7.1.2.2)
- Calculate the mean, the standard deviation and the coefficient of variation (%) for these concentrations
- Calculate the trueness as bias (%)

To determine if the bias is significantly different from 0 a t-test has to be performed, accepting/rejecting the “0” hypothesis.

$$t_{calc} = \frac{bias - 0}{s_{bias} \cdot \sqrt{\frac{1}{n}}} = \frac{bias}{u_{bias}} \leq t_{crit}$$

Where:

bias is the difference between the mean calculated concentration and the certified

reference value,

n is the number of replicates (6 in this case)

s_{bias} is the standard deviation of the six repeated analyses.

If $t_{calc} > t_{crit}$ the hypothesis is rejected and the bias is significantly different than "0". A decision should be taken based on pre-defined acceptability criteria whether the bias is acceptable or not.

Worked example

Samples: One set of 6 analyses per each available CRM has to be carried out (+ blank).
i.e. 7 analyses on day 1 by operator A or B

5.2.7.2.2.2 Reference materials (RM)

As the availability of CRMs is limited, reference value for trueness experiments could also be taken from RM.

Where CRMs are not available, or as an addition to CRMs (if additional concentrations different from the one of the CRM are needed), use may be made of any material sufficiently well characterised for the purpose (a reference material, checked in-house for stability and retained for in-house quality control).

It must be always considered that while insignificant bias may not be proof of zero bias, significant bias on any material remains a cause for investigation.

The calculations and analyses to be performed are the same as for the CRM.

5.2.7.2.2.3 Methods for which organised collaborative studies and proficiency testing results are available

The trueness of the analytical method is examined by participating in a proficiency testing scheme for samples equivalent to the type of sample the method will be used for. The documented trueness only applies to the concentration range and the matrices which are included in the proficiency testing scheme.

If such testing does not exist, smaller comparisons with a few laboratories, (at least 4) after rejection of outliers, can in some cases provide valuable information.

5.2.7.2.2.4 Reference method (RMethod)

In this context, a reference method is an analytical method for the same compound of interest, which has been studied in a collaborative study with good results, and which has an acceptable trueness.

This is a useful option when checking an alternative to, or a modification of, an established standard method already validated and in use in the laboratory..

To do this, samples with 3 different concentration levels (covering the whole range of analysis) have to be analysed at least 6 times, using the two methods.

All values coming from the analyses of the sample according to the reference method and the tested method have to be statistically checked for normality with Shapiro-Wilk Normality test, and Grubbs test for outliers as explained in 5.2.7.1.2.2 and for homogeneity of variance.

To verify whether there is a significant difference between the results obtained for the reference method and those obtained for the method which is to be validated internally, a t-test has to be performed:

A t value is calculated from the 2 mean values obtained from the analyses performed with the 2 different methods as follows:

$$t_{calc} = \frac{\left(\bar{x}_i - x_{ref} \right)}{s_{combined} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq t$$

Where:

$\bar{x}$ is the mean value obtained with the tested method,

$\bar{x}_{ref}$ is the mean value obtained with the reference method,

n is the number of replicate analyses performed for the reference method (6 in this example)

m is the number of replicate analyses performed for the tested method (6 in this example),

s is the combined standard deviation calculated according to the following formula:

$$s = \sqrt{\frac{1}{m+n-2} * [(n-1) * s_x^2 + (m-1) s_{ref}^2]}$$

If the calculated t-value is lower than t critical for P probability level (generally 95%), and m+n-2 degrees of freedom, there is no significant difference between the two sets of results and the tested method can be considered comparable to the reference method and thus the trueness of the reference method can be used as trueness of the tested method.

Worked example

Samples: 6 analyses for each of the 3 concentration levels have to be performed according to the reference method (+ blank).
24 analyses on day 1 by operator A or B

Calibration curve: a new calibration curve has to be established for the reference method (5 concentration + blank, analyses in duplicate)

12 analyses in day 1 by operator A or B)

36 analyses on 3 sample in day 1 by operator A or B)

5.2.7.2.2.5 Reference Laboratory

In this case, the samples are sent to another laboratory, preferably one that is accredited for the relevant method, and the same homogeneous samples are analysed in both laboratories.

To assess if there is a significant difference between the results obtained by the two laboratories a t-test has to be performed, as explained in the paragraph above (for reference method).

It must always be considered that while insignificant differences between the two

results may not be a proof that the method completely behaves like a fully validated one, significant bias on any material remains a cause for investigation.

No new analyses have to be performed by the laboratory.

5.2.7.2.2.6 Spiking/Recovery - Method for which certified reference materials, reference methods or collaborative study or proficiency testing results are not available

In the absence of reference materials or collaborative studies or PT results, bias can be investigated by spiking and recovery. A typical test material is analysed by the method under validation both in its original state and after the addition (spiking) of a known mass of the analyte to the test portion. The difference between the two results as a proportion of the mass added is called the recovery.

To determine the recovery, experiments using spiked blank matrix should be carried out as follows:

- from the reproducibility study on 3 different spiking concentrations (low, medium and high) from the working range analysed in 3 different batches under reproducibility conditions, a calculation is performed on the mean concentration for each concentration investigated,
- using the equation below, the mean recovery is calculated for each spiking concentration level:

$$R_i\% = \frac{(\bar{x}_i - x_0) * 100}{x_{i\,spike}}$$

Where:

$\bar{x}_i$ is the mean concentration determined in the spiked sample;

x_0 is the concentration of the analyte in the sample before spiking, normally blank sample, or real sample for the method of standard addition;

x_{spike} is the spiked concentration

Strictly, recovery studies as described here only assess bias due to effects operating on the added analyte; the same effects do not necessarily apply to the same extent to the native analyte, and additional effects may also apply to the native analyte. Spiking/recovery studies are accordingly very strongly subject to the observation that while good recovery is not a guarantee of trueness; poor recovery is certainly an indication of lack of trueness.

No new analyses are required

The samples, calibration curves and relative analyses are the same as for reproducibility assessment. No additional analyses have to be performed.

A significance test is used to determine whether the mean recovery for the working range is significantly different from 1.0. The test statistic t is calculated using the following equation:

$$t_{calc} = \frac{\left|1-(\bar{R}/100)\right|}{s_{Rec}\cdot\sqrt{\frac{1}{n}}} = \frac{\left|1-(\bar{R}/100)\right|}{u_{Rec}} \leq t_{crit}$$

Where:

$\overline{R} = (\sum R_i)/n$ is the mean recovery for each of the 3 concentration levels

$u_{\bar{R}} = \frac{s}{\sqrt{n}}$ is the uncertainty of the mean recovery value

s is the standard deviation of the mean recovery in the working range

This value of $t_{(calculated)}$ is compared with the 2-tailed critical value t_{crit}, for kp -1 degrees of freedom at 95% confidence (where kp is the number of results used to estimate $\overline{R}$). If it is greater than or equal to the critical value t_{crit} then $\overline{R}$ is significantly different from 1.

Moreover, in case the calculated t value is higher than the t critical valuea decision should be made, based on pre-defined acceptability criteria, as to whether the recovery is acceptable or not.

5.2.7.2.3 Acceptability criteria for trueness

In cases when bias is significantly different from “0” and/or recovery is significantly different from “1 or 100%”, then an assessment should be performed to establish whether or not the bias/recovery are acceptable according to the fixed pre-defined maximum acceptable criteria

For the field of FCM the following maximum acceptable criteria for recovery are laid down based on convention:

Concentration	Mean recovery [%]
≤ 0.01 ppm (≤10 ppb)	40-120
0.1-0.01 ppm (100-10 ppb)	60-110
≥ 0.1 ppm (≥ 100 ppb)	80-110

Table 8: Maximum acceptable recovery of quantitative methods according to CEN TS 15356, expanded for lower level

In Table 9 the maximum acceptable values of bias, according to 2002/657/EC [8], are presented:

Concentration	Bias range %
≤ 1 ppb	- 50 to 20
1-10 ppb	- 30 to 10
≥ 10 ppb	- 20 to 10

Table 9: Maximum acceptable bias of quantitative methods according to [8]

Action should be taken in case a significantly different recovery is accepted based on

maximum acceptable recovery in Table 8. The laboratory should decide whether to report the results corrected for recovery or to include the low recovery in the estimation of the uncertainty of the method. In any case it is fundamental to state clearly if the results are corrected for the recovery or if they are not.

In cases when significant bias is accepted based on the above maximum acceptable bias in Table 9, it should be taken into consideration when calculating the uncertainty of the method and of the reported result.

5.2.8 Uncertainty

Total (expanded) uncertainty of measurement (U) is a parameter associated with the result of a measurement that characterises the dispersion of the values and could be regarded as a single expression of the accuracy of the analytical method.

X= x ±U

where:

X= real value

x= measured value

U= expanded uncertainty

Expanded uncertainty (U) can then be calculated by multiplying the combined standard uncertainty (u_c) by a factor k that associates to the uncertainty a determined level of confidence.

$$U = k \cdot u_c$$

k value can be obtained in Table 2 in Annex 1, according to a confidence of 95% and to the respective effective degrees of freedom ν_{eff}, as explained later in this chapter.

A simplification is given by using k=2, which gives a level of confidence of about 95%.

The measurement uncertainty of an analytical result may be estimated by a number of procedures, notably those described by [28], [2] and Horwitz (which proposes a general correlation between uncertainty and concentration level). These documents recommend procedures based on a component-by-component approach [28], method validation data ([2], ISO), internal quality control data ([2], ISO), and proficiency test data ([2], ISO). The need to undertake an estimation of the measurement uncertainty using the GUM [28] component-by-component approach is not necessary if the other forms of data are available and used to estimate the uncertainty. The Horwitz approach is used when no other data are available (for example in the beginning of a validation of a new method) and provides only an estimation of the real uncertainty. It can be used to verify if the method can be suitable for the purpose, but it is not sufficient for the determination of the real uncertainty.

In the present chapter a detailed description of all these approaches will be presented, following a case by case scheme.

5.2.8.1 Horwitz approach

When none of the other approaches are possible, the following formula can be used to estimate the uncertainty:

$U\,(\%)_{predicted} = 2^{(1-0.5\log C)} = 2*C^{(-0.5\log 2)} = 2*C^{-0.15}$

Where:

C is the mass fraction expressed as a power (exponent) of 10 (e.g. 1 mg/g = 10-3).

5.2.8.2 GUM (component-by-component) approach [28]

When no proficiency testing scheme, certified reference materials or internal quality control are available and where no real reproducibility can be assessed, a calculation of the uncertainty of the results is not possible. In such cases all possible sources of uncertainty (combined uncertainty) not included in repeatability S, have to be taken into account, considering a determined level of confidence. Thus the final expanded uncertainty will be two times the combined uncertainty value that can be calculated by the combination of the uncertainty coming from the calculation of reproducibility and all other source of uncertainty:

$U_{exp} = 2u_c = 2\sqrt{s_{répeat}^{\ 2} + u_{typeB}^{\ 2}}$

Where:

U_{exp} is the expanded uncertainty;

U_c is the combined uncertainty;

S_{repeat} is the standard deviation obtained from at least 6 samples analyses (those from repeatability study) integrating all uncertainty sources due to the sample treatment;

$u_{type\ B}$ is given by the combination of all possible sources of uncertainty which are not included in s_{repeat}: $\sqrt{\sum u_i^{\ 2}}$

Combined relative uncertainty (u_c) is the uncertainty deriving from the combination of all associated uncertainties (u) of all the different stages of the method, such as uncertainty on volumes (such as those related to pipettes, volumetric flasks, etc.), weighing, measures and the uncertainty due to errors deriving from the calibration curve. For chemical measures it can be calculated by a general formula:

$$\dot{u}_c(\bar{y}) = \frac{u_c(\bar{y})}{\bar{y}} = \sqrt{\left(\frac{u(\bar{y}_A)}{\bar{y}}\right)^2 + \left(\frac{u(\bar{c}_B)}{\bar{c}}\right)^2 + \left(\frac{u(\bar{v}_B)}{\bar{v}}\right)^2 + \left(\frac{u(\bar{p}_B)}{\bar{p}}\right)^2 + \left(\frac{u(\bar{R}_B)}{\bar{R}}\right)^2 + \left(\frac{u(\bar{M}_B)}{\bar{M}}\right)^2}$$

Where:

$\frac{u(\bar{y}_A)}{\bar{y}} = \frac{s}{\bar{x}\cdot\sqrt{n}} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}}{\bar{x}\cdot\sqrt{n}}$ is the relative uncertainty due to repeatability, S_{repeat}

Where:

s is the Standard deviation, an estimate of the dispersion of the data (n) around the mean value $\bar{x}$,

$\bar{x}$ is the mean value,

x_i is the measured values,

n is the number of analyses.

$$\frac{u\left(\bar{c}_B\right)}{\bar{c}_s} = \frac{\sqrt{\bar{s}}}{\bar{c}_s \cdot \sqrt{q}}.$$

is the uncertainty due to errors on calibration curves.

Where:

$\bar{c}_s$ is the mean value of all standard concentration values of for each point of the calibration curve

q is the number of samples for repeatability analyses.

$\bar{s} = \frac{\sum s_i}{q}$ is the mean value of the squared values of the q samples of the contributions linked to repeatability and calibration curves errors

Where:

$$s_i = \left(\frac{\sqrt{\frac{\sum_{i=1}^{n}\left[y_i - \hat{y}_i\right]^2}{n-2}}}{b} \cdot \sqrt{\left(\frac{1}{m} + \frac{1}{n} + \frac{\left(y_j - \bar{y}\right)^2}{b^2 \sum_{i=1}^{n}\left(c_i - \bar{c}_c\right)^2} \right)} \right)$$

Where:

y_j is the value given by the instrument for each measure of the repeatability assessment (peak area)

y_i is the value given by the instrument for each measure of all points of the calibration curve (peak area)

$\bar{y}$ is the mean value of all values given by the instrument for each measure of all points of the calibration curve (peak area)

$\hat{y}_i =$ a+bc$_i$ is the value calculated with calibration curves parameters for each point of calibration (peak area)

c_i is the concentration of standards for each point of the calibration curve

$\bar{c}_c$ is the mean value of all standard concentration values of for each point of the calibration curve

b is the angular coefficient

n is the number of points of the calibration curve

m is the number of analyses for each sample (repetitions) for repeatability study.

$$\frac{u\left(\bar{v}_B\right)}{\bar{v}} = \frac{\sqrt{{u_{vs}}^2 + {u_{vr}}^2 + {u_{v\Delta T}}^2}}{\bar{v}}$$

is the uncertainty due to volumes, the used volume being $\bar{v}$.

Where:

$u_{vs} = \frac{u_{vs}}{\bar{v} \cdot \sqrt{3}}$ The first term is the uncertainty declared by the supplier.

u_S is the value taken from manufacturer certificate.

$u_{vr} = \frac{s_v}{\bar{v} \cdot \sqrt{n_v}}$ = uncertainty due to repeatability in filling.

s_V is the standard deviation due to variation in filling and it is calculated performing a repeatability experiment on n_v fillings and weightings.

$u_{v\Delta T} = \frac{C_e \cdot \bar{v} \cdot \Delta T}{\bar{v}}$ is the uncertainty due to the difference in temperature between the moment when the volumetric glassware was calibrated by the supplier and the moment it is used. C_e is the coefficient of volume expansion and □T is the difference in temperature between calibration of glassware and analyses.

$\frac{u(\bar{p}_B)}{\bar{p}} = \frac{1-p}{\sqrt{3}}$ is the uncertainty due to degree of purity of standards, where p is the purity value as from the certificate of standard material.

$\frac{u(\bar{R}_A)}{\bar{R}} = \frac{s_{rec}}{\sqrt{n_r}}$ is the uncertainty due to recovery, calculated on n_r analyses of spiked samples, where s_r is the standard deviation of n_r analyses performed to calculate recovery. A significance test is used to determine whether the mean recovery is significantly different from 1.0 (or 100%).

$\frac{u(\overline{M}_B)}{\overline{M}} = \frac{\sqrt{\sum\left(\frac{u_{Mi}}{\sqrt{3}}\right)^2}}{\overline{M}}$ is the relative uncertainty due to weighing. One contribution must be considered for every utilisation of the balance (also for tare weigh).

$\overline{M}$ is the mean value of the weightings of the samples for repeatability assessment.

$u_{Mi=} \sqrt{2*(u_{MR})^2 + (u_{ME})^2 + (u_{M\Delta T})^2}$ is the contribution given by each use of the balance

Where:

u_{MR} is the uncertainty due to repeatability, twice because of the tare weight.

u_{ME} is the uncertainty due to eccentricity (from calibration certificate)

$u_{M□T}$ is the uncertainty due to temperature variations given by:

$$\left(\frac{K \cdot M \cdot \Delta T}{\sqrt{3}}\right)^2$$

Where:

K is a constant reported in the balance certificate,

M is the weighted mass,

ΔT is the difference in temperature between the calibration and the weighting.

The general expression of uncertainty has to be adapted to the considered method, eventually adding all necessary additional terms for the calculation of all components of uncertainty or deleting the terms that are not necessary as already included in other estimations.

The numbers of freedom degreeș ν_{eff} can be calculated by:

$$\nu_{eff} = \frac{\left[\dot{u}_c\left(\bar{y}\right)\right]^4}{\left[\frac{\left(\dot{u}\left(\bar{x}_r\right)\right)^4}{\nu_r}\right] + \left[\frac{\left(\dot{u}\left(\bar{x}_m\right)\right)^4}{\nu_c}\right]}$$

Where:

ν_r is the number of analyses performed to calculate repeatability

ν_c is the number of calibration curve points

$\dot{u}\left(\bar{x}_r\right)$ is the relative uncertainty on repeatability

$\dot{u}\left(\bar{x}_m\right)$ is the relative uncertainty on measures

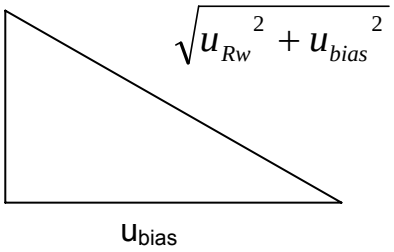
$\dot{u}_c\left(\bar{y}\right)$is the relative combined uncertainty.

5.2.8.3 Calculation of uncertainty when method validation data, internal quality control data and proficiency test data are available

With well-planned validation studies it is often possible to assess the uncertainty of a large part of the existing elements of uncertainty. The general approach for this requires estimation of the random and systematic errors of the method in use. These provide estimates of within- and between-laboratory components of variance, together with an estimate of uncertainty associated with the trueness of the method. Other sources of uncertainty should be quantified in separate experiments or, if possible, by checking if there is enough information in the given specification or literature data. Data could be collected from analyses of CRMs, participation in collaborative studies, participation in proficiency testing schemes, internal quality control tests, results from duplicate samples, recovery from spiking etc.

The combined standard uncertainty u_c is calculated as the square root of the quadratic sum of the random component u_{Rw} and the systematic component u_{bias}:

$u_c = \sqrt{u_{Rw}^2 + u_{bias}^2}$



The procedure for the determination of measurement uncertainty essentially consists of the steps shown in Figure 9.

All data used for the calculations described in Figure 9 are the same obtained for the calculation of accuracy

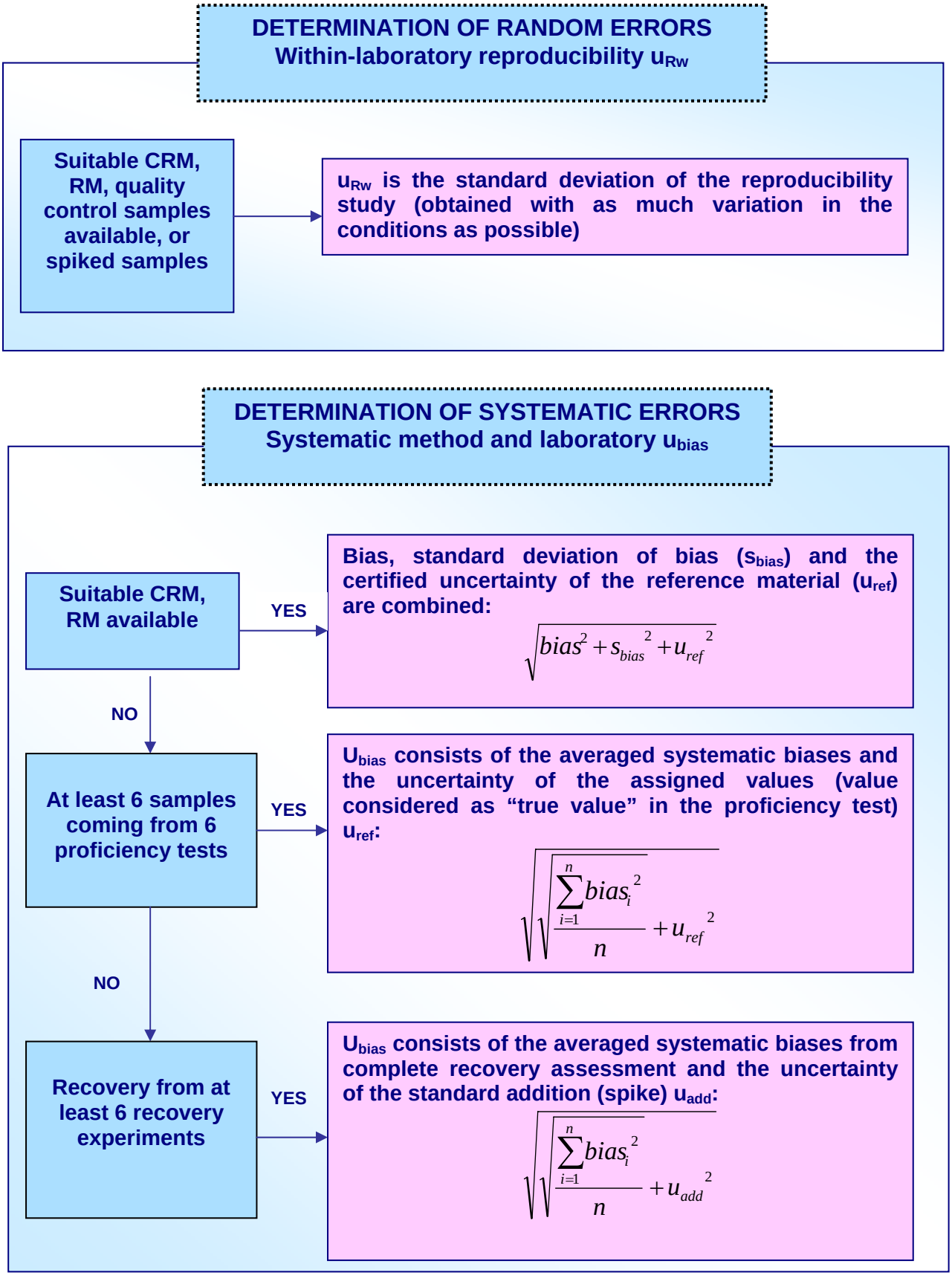


Figure 9: Estimation of the random and systematic component of measurement uncertainty.

5.3 Reporting of results

The reporting of the analytical results depends on how the limits or restrictions specified in the legislation are expressed. Thus, the number of significant figures used when expressing an analytical result has to be consistent with what is expressed in the legislation.

In cases where legislation already provides clear guidance on the number of significant figures to be specified (e.g. when maximum levels are expressed as 2.00, 15.0, or 0.040) then the analyst should report the result to the same number of significant figures as indicated in the specification. In other cases, and certainly in cases where it is appropriate for the precision of the result, the analyst should use one significant figure more than indicated in the specification (assuming that the analyst is using an appropriate method).

The analytical result has to be reported as x +/- U, where x is the analytical result (the best estimate of the true value) and U is the expanded measurement uncertainty. The result is reported with uncertainty of measurement when it is relevant to the validity or application of the test results, when a customer's instruction so requires, or when the uncertainty affects compliance to a specification limit [21].

5.4 Interpretation of results

It is recommended to use the measurement uncertainty when assessing compliance with a specification.

In practice, if the legislation specifies a certain limit value, the analyst has to estimate the measurement uncertainty at that level. The value obtained by subtracting the uncertainty from the reported result is used to assess compliance. Only if this value is greater than the limit given in the legislation is it certain "beyond reasonable doubt" that the sample concentration of the analyte is greater than the specified limit.

5.5 Validation/verification documentation

Once the validation study is performed, the final validation or verification documentation should include:

- Records on planning and preparation of the validation study.
- All raw data or a reference to where the raw data may be found.
- An accurate specification of which parameters have been examined. If all of the parameters have not been examined the reason(s) must be reported.
- Obtained results and how these have been calculated.
- It is especially important to include on which matrices and at which concentrations the specific precision, trueness, detection limits and quantification limits have been determined.
- Evaluation of the obtained results in relation to the validation plan or verification plan.

- An unambiguous conclusion concerning which analytical tasks the examined method is suitable for.
- If the method is found to be unsuitable for some matrices or concentrations this must be evident in the report.

Thus the validation/verification report should include:

- Title of method
- Analyte
- Matrix
- Description of method
- Brief description of study (plan)
- Results of study (reporting results for each single parameter, their comparison with the eventual assessment criteria, etc.)
- Conclusions

It is essential to record exactly how and on which materials the repeatability or internal reproducibility (reference material, control material, authentic samples or synthetic solutions) has been determined. If the analytical method being studied is to be used for a large concentration range, the analyte precision should be estimated at several concentration levels, usually low, medium and high. It is recommended that the estimation of precision is repeated in part or in full during the validation work.

5.6 Monitoring of the validated method

When a validated/verified method is taken into routine use in a laboratory, it is important that a competent analyst is made responsible for monitoring that the method continuously performs in accordance with the results obtained in the validation/verification.

The trueness and precision of the method must be monitored continuously. The results obtained must be in accordance with the results of the validation or verification study. How often such control should be carried out, will be determined by the responsible analyst and must be thoroughly documented, for example in the internal analytical user manuals.

Changes in the experimental conditions for a method can cause a modification of the function of the method. In such cases it may be necessary to carry out a new complete validation.

6 “STANDARD LEVEL” VALIDATION SCHEME FOR FCM METHODS

All the theoretical considerations and description of the different parameters are already extensively presented in the previous chapter. Thus the present chapter will concentrate mostly on the practical part of the analytical work that has to be performed by the laboratory.

All information reported in paragraphs 5.3, 5.4, 5.5, 5.6 should also be applied also for the standard FCM validation scheme.

6.1 Selectivity/specificity

The assessment of selectivity can be based on checking for potential interferences at the legislative limit. This means that the matrix has to be analysed at the legislative limit and the presence of any substance that can interfere with the analysis of the target analyte has to be assessed. No specific analyses are foreseen for the assessment of this parameter, as all necessary data can be obtained from the recovery study and the calibration curves.

6.2 Ruggedness

It is considered that ruggedness should be performed during method development and therefore it is not necessary for this validation scheme. However, if new types of analytical equipment are introduced ruggedness testing is needed.

6.3 Calibration range, assessment of calibration function

6.3.1 Calibration range

According to 5.2.3.1 for full validation

6.3.2 Basic of calibration and quantification

It can be performed according to 5.2.3.2 for full validation with some small changes in the experimental plan.

In order to simplify the procedure, when possible, the methods are divided into two main categories (for inorganic and organic analyses). Inorganic and organic analyses present different requirements and are assessed in two different ways:

- Inorganic analyses: 3 concentration levels plus one blank have to be analysed and the procedure has to be repeated 3 times (12 analyses in total).
- Organic analyses: 5 concentration levels plus one blank have to be analysed and the procedure has to be repeated 3 times (18 analyses in total).

6.3.3 Assessment of the linearity model

For the present validation scheme the calibration function has to be assessed initially by visual inspection of the plotted data generating a calibration curve and then by statistical evaluation.

The correlation coefficient, *r*, is not a measure of linearity. It indicates the extent of the relationship between the instrument response and analyte concentration.

For the assessment of the regression model for the calibration function it is important to carry out first the visual inspection of the plot of residuals (difference between the measured value and the expected value) versus the corresponding concentration (x) values.

Using the data from the calibration curves of the reproducibility assessment, estimation and examination of the residuals has to be performed and the requirements in paragraph 5.2.3.3 should be met.

A statistical treatment of data is necessary. For this purpose ANOVA lack of fit or Mandel's test or another equivalent test has to be performed (see paragraph 5.2.3.4 for all details regarding the tests).

In some cases where the calibration range is very wide it is important to establish the homogeneity of the variance of the method across the working range according to the procedure described in 5.2.3.5.

As related to "homoscedasticity" a distinction between spectrometric and chromatographic methods has to be made:

- for spectrometric methods the test for homogeneity of the variance ("homoscedasticity") is suggested only if necessary (in cases with a very large working range);
- for chromatographic methods, as the working range is usually narrow, calibration without checking for "homoscedasticity" is accepted and calibration without weighting is suggested .

6.4 Working concentration range

As described in paragraph 5.2.4.

No specific analyses have to be performed for the assessment of the working range, as data coming from precision experiments can be used.

6.5 LOD

The calculation of LOD is not necessary, except for the cases when LL=ND

6.6 LOQ

LOQ is calculated only when necessary

A distinction between spectrometric and chromatographic methods has to be made:

- For chromatographic methods estimation of the LOQ should be based on the lowest level of the calibration curve and should be equal to the concentration at which the ratio between the signal to noise (S/N) is at least 6.
- For spectrometric methods estimation of LOQ should be based on the blank and should be equal to the average response of the blank + at least 6 times the standard deviation of the blank or the lowest point if the blank does not give a signal.

It is necessary to perform a verification of the quality of the value estimated as LOQ following paragraph 5.2.6.1.

All data necessary for the calculation of LOQ are those collected for the assessment of precision (lowest point). Thus no new analyses have to be performed for the calculation of LOQ.

6.7 Precision

Repeatability and within-laboratory reproducibility can be assessed together, based on the following schemes.

The choice of the number of replicates to be analysed each day is due to the necessity to reach a number of degrees of freedom (minimum 6) that can ensure the precision of the procedure. To this purpose, any combination of analyses presented in Table 10, giving a degree of freedom not less then 6 can be used for each of the 3 levels of concentration within the working range (0.2*LL, LL, 2*LL, or 5*LL when reduction factors are applied).

ANOVA can be used to calculate both repeatability and reproducibility: repeatability is represented by the within batch standard deviation and reproducibility is represented by square root of the sum of the squares of the within batch standard deviation and the between batches standard deviation.

Within-laboratory reproducibility standard deviation must be determined by at least 6 degrees of freedom.

Number of samples in each batch	Number of batches	Degrees of freedom for repeatability standard deviation	Degrees of freedom for internal laboratory reproducibility standard deviation
7	1	6	Not determined
4	2	6	7
3	3	6	8
2	6	6	11
n	m	(n-1)*m	n*m-1

Table 10: Number of degrees of freedom obtained by combinations of number of batches and number of samples in each batch.

As an alternative to ANOVA data treatment, the following scheme can be used:

Day 1: at least 7 replicate samples at the legislative level (for repeatability assessment) and other 2 replicates at 2 other concentration levels in the working range in food/ food simulant (e.g. 0.2*LL and 2*LL) plus the calibration curve should be analysed.

Day 2: at least 3 replicates at 3 concentration levels (0.2*LL, 1*LL and 2*LL) and calibration curves should be analysed.

Day 3: at least 3 replicates at 3 concentration levels (0.2*LL, 1*LL and 2*LL) and calibration curves should be analysed.

Basic statistical evaluation of the data for verification of the presence of anomalous data should be performed by applying Grubbs or Dixon's tests as described in 5.2.7.1.2.2

Calculation of within-laboratory reproducibility as described in 5.2.7.1.3.3.

Acceptability criteria for precision, described in 5.2.7.1.4 should be met.

Worked example without ANOVA

*Samples: A total of 21 analyses had to be performed. First day - 1 concentration (LL) in 7 replicates, two other concentration levels (0.2*LL and 2*LL) in 3 replicates; two other different days the 3 concentration levels (0.2*LL, 1.0*LL and 2.0*LL) in 3 replicates by different operator.*
=> 13 analyses in day 1, 9 analyses in day 2, 9 analyses in day 3

3 calibration curves:
For inorganic analyses: A total of 12 analyses had to be performed (3 times 3 concentrations+ blank) in different days by three different operators

For organic analyses: A total of 18-21 analyses had to be performed (3 times 5-6 concentrations+ blank) in different days by different operators.

⇨ ***33 analyses by different operators in 3 days for inorganic analyses***
⇨ ***39-42 analyses by different operators in 3 days for organic analyses***

Worked example with ANOVA

The scheme of analyses is the same as for the previous case, but the numbers vary according to the combination chosen from table 6 at 3 different concentration levels.

6.8 Trueness

Trueness is calculated on the basis of bias or recovery using whatever samples/data are available (CRM, RM, proficiency testing results, spiked in- house materials).

In case of availability of CRM or RM, the bias should be estimated based on 3 repeated analyses of these materials, performed together with the samples for recovery study, thus no new calibration curve is needed as those for the reproducibility assessment can be used.

For spiked samples 3 replicates of spiking of the analyte in the sample for 3 levels of concentration within the working range (0.2*LL, LL, 2*LL, or 5*LL when a reduction factor is applied) have to be performed.

Also in this case data collected for precision study can be used.

Acceptability criteria for trueness, described in 5.2.7.2.3 should be met.

6.9 Uncertainty

Uncertainty can be estimated as explained in paragraph 5.2.8 and in Figure 9.

7 “BASIC LEVEL” VALIDATION SCHEME FOR FCM METHODS

All the theoretical considerations and description of the different parameters are already extensively presented in the previous chapters 5 and 6. Thus the present chapter will concentrate mostly on the practical part of the analytical work that has to be performed by the laboratory.

All information reported in paragraphs 5.3, 5.4, 5.5 and 5.6 should also be applied also to the basic validation scheme.

7.1 Selectivity/specificity

The assessment of selectivity can be based on checking for potential interferences at the legislative limit. This means that the matrix has to be analysed at the legislative limit and the presence of any substance that can interfere with the analysis of the target analyte has to be assessed. No specific analyses are foreseen for the assessment of this parameter, as all necessary data can be obtained from recovery study and calibration curves.

7.2 Ruggedness

Ruggedness is not considered not necessary for this validation scheme.

7.3 Assessment of calibration function

7.3.1 Calibration range

According to 5.2.3.1 for full validation

7.3.2 Basic of calibration and quantification

According to 6.3.2 for standard validation. The assessment should be done on 1 calibration curve prepared in replicate as described in Figure 3, the flow chart for basic level validation.

7.3.3 Assessment of the linearity model

For the assessment of the regression model for the calibration function it is important to carry out first the visual inspection of the residual plot. There should be no curvature in the residuals plot as described in paragraph 5.2.3.4.2 and as shown in Figures 6 and 7;

No statistic treatment of data is necessary.

As related to the “homoscedasticity” distinction between spectrometric and chromatographic methods has to be made:

- for spectrometric methods the test for homogeneity of the variance (“homoscedasticity”) is suggested only if necessary (in cases with a very large working range);
- for chromatographic methods as the working range is usually narrow. calibration without checking for “homoscedasticity” is accepted and calibration without weighting is suggested.

7.4 Working concentration range

The working concentration range is the range in which the method is validated and which gives an acceptable trueness and precision

As the basic level of method validation is focusing only on the LL, a narrow working range around that LL is considered to be acceptable.

7.5 LOD

The calculation of LOD is not necessary except for the cases when LL=ND.

7.6 LOQ

A distinction between spectrometric and chromatographic methods has to be made:

- For chromatographic methods estimation of LOQ should be based on the lowest level of the calibration curve and should be equal to the concentration at which the ratio between the signals to noise (S/N) is at least 6.
- For spectrometric methods estimation of LOQ should be based on the blank and should be equal to the average response of the blank + at least 6 times the standard deviation of the blank or the lowest point if the blank does not give a signal.

All data necessary for the calculation of the LOQ are those collected for the assessment of precision (lowest point). Thus no new analyses have to be performed for the calculation of LOQ.

7.7 Precision

Repeatability and within-laboratory reproducibility can be assessed, based on the following schemes.

The choice of the number of replicate samples to be analysed each day is due to the necessity to reach a number of degrees of freedom that can ensure the precision of the procedure. To this purpose, Table 6 can be used (as described in chapter 6).

ANOVA can be used to calculate both repeatability and reproducibility: repeatability is represented by the within batch standard deviation and reproducibility is represented by square root of the sum of the squares of within batch standard deviation and between batches standard deviation.

Within-laboratory reproducibility standard deviation must be determined by at least 6 degrees of freedom.

For laboratories not familiar with ANOVA data treatment, the following scheme can be used:

For repeatability: at least 5 replicated samples for legislative level in the working range in food/food simulant, plus calibration curve.

For reproducibility: 3 replicates plus calibration curves in another 2 different days (repeatability analyses can be considered to be day 1)

Basic statistical evaluation of the data for verification of the presence of anomalous data should be performed by applying Grubbs or Dixon's tests as described in 5.2.7.1.2.2.

Acceptability criteria for precision, described in 5.2.7.1.4 should be met.

Worked example without ANOVA

Samples: A total of 11 analyses had to be performed (1 concentration - 5 times the first day, 3 times on day 2 and 3 times on day 3) by different operator.
=> 5 analyses on day 1, 3 analyses on day 2, 3 analyses on day 3

3 calibration curve:
For inorganic analyses: A total of 12 analyses had to be performed (3 times 3 concentrations+ blank) by different operators in different days.

For organic analyses: A total of 18 analyses had to be performed (3 times 5 concentrations+ blank) by one operator.

⇨ ***23 analyses by different operators in 3 days for inorganic analyses***
⇨ ***29 analyses by different operators in 3 days for organic analyses***

7.8 Trueness

Trueness is calculated on the basis of bias or recovery using whatever samples/data are available (CRM, RM, proficiency testing results, spiked in- house materials).

In case of availability of CRM or RM, the bias should be estimated based on 3 repeated analyses of these materials, performed together with the samples for recovery study, thus no new calibration curve is needed as those for the reproducibility assessment can be used.

For the recovery study spiking the sample at the legislative limit, three replicates are considered sufficient. Also in this case data collected for the reproducibility assessment can be used.

Acceptability criteria for trueness, described in 5.2.7.2.3 should be met.

7.9 Uncertainty

Uncertainty can be estimated by the standard deviation calculated on within-laboratory reproducibility.

8 REFERENCES

1. AOAC Peer Verified methods Program, Manual on policies and procedures, Arlington, VA, Nov.1993.

2. Eurachem/CITAC, 2000, Quantifying uncertainty in analytical measurement (2nd edition).

3. Draft SANCO/1805/2000 rev. 1 Laying down performance criteria for the analytical methods to be used for certain substances and residues thereof in live animals and animal products according to Council Directive 96/23/EC.

4. ISO Standard 5725 1-6: Accuracy (trueness and precision) for measurement methods and results.

5. Statistical Manual of the AOAC, Association of Official Analytical Chemists, Arlington, VA, USA, 1975.

6. Manuale N. 179/1 UNICHIM - Linee guida per la validazione di metodi analitici nei laboratori chimici - valutazione della precisione (ripetibilità stretta) di un metodo analitico eseguito in un unico laboratorio da un solo operatore su di un unico strumento in un breve intervallo di tempoLinea guida per validazione di metodi analitici nei laboratori chimici. Edz. 2001

7. AOAC Guidelines for Single Laboratory Validation of Chemical Methods, 2002.

8. Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (2002/657/EC).

9. Draft ISO TS 21748 – “Guide to the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation”, Geneva, 2003

10. DIN 38 402 (Teil 51): Deutsche Einheitsverfahren zur Wasser-, Abwasser und Schlammuntersuchung. Allgemeine Angaben (Gruppe A) Kalibrierung von Analysenverfahren, Auswertung von Analysenergebnissen und lineare Kalibrierfunktionen für die Bestimmung von Verfahrenskenngrößen (A51) (1986) 5.1.3 Linearitätstest.

11. ISO standard 11095:1996: Linear calibration using reference material

12. ICH Q2B Validation of analytical procedures: text and methodology, Q2(R1), November 2005.

13. CIPAC Guidelines for Collaborative Study Procedures for Assessment of Performance of Analytical Methods, 1989.

14. Food and Drug Administration, Reviewer Guidance on "Validation of Chromatographic Methods", 1994.

15. IUPAC Compendium of Chemical Terminology, 2nd Edition (1997)

16. CEN/TC194/SC1/WG1 N260: CEN Report on the validation and interpretation of the analytical methods, migration testing and analytical data for materials and articles in contact with food, Part 1, General considerations. (CEN TR 15330)

17. Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing; Thompson, M., Analyst, 125, 385-386 (2000).

18. “Statistics and Chemometrics for Analytical Chemistry” by J. Miller, J. Miller , Pearson Education Limited, 5th edition, 2005, chapter 3.9

19. “Fundamental Concepts in the Design of Experiments”, C R. Hicks, K.V. Turner

20. ISO Standard 78-2: Chemistry – Layouts for Standards – Part 2: Methods of Chemical Analysis (Second Edition, 1999).

21. ISO/IEC 17025:2005: General requirements for the competence of testing and calibration laboratories.

22. 2002/72/EC: Commission Directive of 6 August 2002 relating to plastic materials and articles intended to come into contact with foodstuffs.

23. 82/711/EEC: Council Directive of 18 October 1982 laying down the basic rules necessary for testing migration of the constituents of plastic materials and articles intended to come into contact with foodstuffs.

24. 85/572/EEC: Council Directive of 19 December 1985 laying down the list of simulants to be used for testing migration of constituents from plastic materials and articles intended to come into contact with foodstuffs.

25. ISO 8466-1:1990: Water quality -- Calibration and evaluation of analytical methods and estimation of performance characteristics -- Part 1: Statistical evaluation of the linear calibration function.

26. ISO 8466-2:2001: Water quality -- Calibration and evaluation of analytical methods and estimation of performance characteristics -- Part 2: Calibration strategy for non-linear second-order calibration functions.

27. Manuale N. 179/2 – UNICHIM - Linee guida per la validazione di metodi analitici nei laboratori chimici - valutazione della precisione (ripetibilità) di un metodo analitico eseguito in un unico laboratorio con più operatori/strumenti. Edz. 1995.

28. ISO/IEC Guide 98:1995 - Guide to the expression of uncertainty in measurement (GUM).

29. Regulation (EC) No 1935/2004 of the European Parliament and of the Council of 27 October 2004 on materials and articles intended to come into contact with food and repealing Directives 80/590/EEC and 89/109/EEC.

30. Regulation (EC) No 882/2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules.

31. ISO/IEC Guide 43-1:1997 - Proficiency testing by interlaboratory comparisons -- Part 1: Development and operation of proficiency testing schemes.

32. ISO/IEC Guide 43-2:1997 - Proficiency testing by interlaboratory comparisons -- Part 2: Selection and use of proficiency testing schemes by laboratory accreditation bodies.

33. Commission Directive 93/8/EEC of 15 March 1993 amending Council Directive 82/711/EEC laying down the basic rules necessary for testing migration of constituents of plastic materials and articles intended to come into contact with foodstuffs.

34. Commission Directive 97/48/EC of 29 July 1997 amending for the second time Council Directive 82/711/EEC laying down the basic rules necessary for testing migration of the constituents of plastic materials and articles intended to come into contact with foodstuffs.

35. DIN 32645:2008-11 - Chemical analysis - Decision limit, detection limit and determination limit under repeatability conditions - Terms, methods, evaluation.

ANNEX 1

TABLES FOR STATISTICAL TESTS

		Numerator degrees of freedom					
		1	2	3	4	5	6
Denominator degrees of freedom	1	4052.2	4099.5	5403.4	5624.6	5763.6	5859.0
	2	98.50	99.00	99.17	99.25	99.30	99.33
	3	34.12	30.82	29.46	28.71	28.24	27.91
	4	21.20	18.00	16.69	15.98	15.52	15.21
	5	16.26	13.27	12.06	11.39	10.97	10.67
	6	13.75	10.92	9.78	9.15	8.75	8.47
	7	12.25	9.55	8.45	7.85	7.46	7.19
	8	11.26	8.65	7.59	7.01	6.63	6.37
	9	10.56	8.02	6.99	6.42	6.06	5.80
	10	10.04	7.56	6.55	5.99	5.64	5.39
	11	9.65	7.21	6.22	5.67	5.32	5.07
	12	9.33	6.93	5.95	5.41	5.06	4.82
	13	9.07	6.70	5.74	5.21	4.86	4.62
	14	8.86	6.51	5.56	5.04	4.69	4.46
	15	8.68	6.36	5.42	4.89	4.56	4.32
	16	8.53	6.23	5.29	4.72	4.44	4.20

Table 1a: F of Fisher values for 1% risk

		Numerator degrees of freedom					
		1	2	3	4	5	6
Deniminator degrees of freedom	1	166.5	199.5	215.7	224.6	230.2	234.0
	2	18.5	19.0	19.2	19.2	19.3	19.3
	3	10.1	9.55	9.28	9.12	9.01	8.94
	4	7.71	6.94	6.59	6.39	6.26	6.16
	5	6.61	5.79	5.41	5.19	5.05	4.95
	6	5.95	5.14	4.76	4.53	4.39	4.28
	7	5.59	4.74	4.35	4.12	3.97	3.87
	8	5.32	4.46	4.07	3.84	3.69	3.58
	9	5.12	4.26	3.86	3.63	3.48	3.37
	10	4.96	4.10	3.71	3.48	3.33	3.22
	11	4.84	3.98	3.59	3.36	3.20	3.09
	12	4.75	3.89	3.49	3.26	3.11	3.00
	13	4.67	3.81	3.41	3.18	3.03	2.92
	14	4.60	3.74	3.34	3.11	2.96	2.85
	15	4.54	3.68	3.29	3.06	2.90	2.79
	16	4.49	3.63	3.24	3.01	2.85	2.74

Table 1b: F of Fisher values for 5% risk

Effective degrees of freedom (□eff)	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Table 2: Factor k values (k=tp, where tp is the Student's t), for different confidence levels

ANNEX 2

CONFIRMATORY CRITERIA FOR CHROMATOGRAPHIC METHODS (GC, LC)

Methods based only on chromatographic analysis without the use of mass spectrometric detection are not suitable on their own for use as confirmatory methods. However, if a single technique lacks sufficient specificity, the desired specificity shall be achieved by analytical procedures consisting of suitable combinations of clean-up, chromatographic separation(s) and mass spectrometric detection.

- The retention time of the analyte shall be the same as that of the calibration standard in the appropriate matrix, within a margin of ± 2.5% (GC) and ± 5%.(LC).

- The ratio of the retention time of the analyte to that of the internal standard, i.e. the relative retention time of the analyte, shall be the same as that of the calibration standard in the appropriate matrix, within a margin of ± 0.5% (GC) and ± 2.5% (LC).

When mass spectrometric detection is used, based on the scanning mode, the following confirmatory criteria can be used:

Full scan: When mass spectrometric determination is performed by the recording of full scan spectra, the presence of all measured diagnostic ions (the molecular ion, characteristic adducts of the molecular ion, characteristic fragment ions and isotope ions) with a relative intensity of more than 10% in the reference spectrum of the calibration standard is obligatory.

SIM: When mass spectrometric determination is performed by single ion monitoring, the molecular ion shall preferably be one of the selected diagnostic ions (the molecular ion, characteristic adducts of the molecular ion, characteristic fragment ions and all their isotope ions). The selected diagnostic ions should not exclusively originate from the same part of the molecule. The signal-to-noise ratio for each diagnostic ion shall be ≥ 3:1.

Full scan and SIM; ***MRM*****:** The relative intensities of the detected ions, expressed as a percentage of the intensity of the most intense ion or transition, shall correspond to those of the calibration standard, either from calibration standard solutions or from spiked samples, at comparable concentrations, measured under the same conditions, within the tolerances shown in Table 1:

Relative intensity (% of base peak)	EI-GC-MS (relative)	CI-GC-MS, GC-MS^n, LC-MS, LC-MS^n (relative)
> 50%	± 10 %	**± 20%**
> 20% to 50%	± 15 %	**± 25%**
> 10% to 20%	± 20 %	**± 30%**
< 10%	**± 50 %**	**± 50%**

Table 1. Maximum permitted tolerance for relative ion intensities using a range of mass spectrometric techniques.

When full-scan UV/VIS detection id used, the maximum absorption in the spectrum of the analyte shall be at the same wavelengths as those of the calibration standard

within a margin determined by the resolution of the detection system. For diode array detection, this is typically within ± 2 nm. The spectrum of the analyte above 220 nm shall, for those parts of the two spectra with a relative absorbance ≥ 10 %, not be visibly different from the spectrum of the calibration standard. This criterion is met when firstly the same maxima are present and secondly when the difference in absorbance between the two spectra is at no point observed greater than 10%. In the case computer-aided library searching and matching are used, the comparison of the spectral data in the test samples to that of the calibration solution has to exceed a critical match factor. This factor shall be determined during the validation process for every analyte on the basis of spectra for which the criteria described above are fulfilled. Variability in the spectra caused by the sample matrix and the detector performance shall be checked.

If absorbance ratios are used the absorbance ratios for the samples should agree to within ±10% of ratios for the standards (at concentrations as close as possible to each other).

LC with UV/VIS detection (single wavelength) is not suitable on its own for use as a confirmatory method.

Annex 3 Quality control and updates

These guidelines will be reviewed at regular intervals. The CRL-NRL Network for FCM is charged with this review and the undertaking of any necessary updates.

Peer –review process

- Drafting: CRL for the CRL-NRL Network with contributions and under advice of a dedicated task force of NRLs.
- Verifying and Consensus; NRL Network
- Approval and Endorsement: European Commission, DG SANCO E3

Updates of text

Versions	Dates	Chapters	Action
1st edition 2009	09.09.2009	all	creation

Quality control box

Written by	Advisement and contributions	Peer-reviewed by	Endorsed by
CRL-FCM Catherine Simoneau	NRL Task Force: NRL-BE (Fabien Bolle, Tina n'Goy), NRL-DE (Oliver Kappenstein), NRL-DK (Jens Petersen), NRL-ES (Juana Bustos), NRL-FR1 (Patrick Sauvegrain), NRL-EL (Timokleia Togkalidou), NRL-IT (Maria Rosaria Milana) NRL-NL (Durk Schakel, Dita Kalsbeek), NRL-PL (Kazimiera Cwiek-Ludwicka), NRL-SI (Viviana Golja), NRL-UK (Emma Bradley).	NRL network See list in annex 2	DG SANCO Annette Schaefer Fabio d'Atri

Annex 4. NRL Network

Member State	Name of NRL
AUSTRIA	Austrian Agency for Health and Food Safety (AGES), Institute für Lebensmitteluntersuchung Wien DI. Marica Peffer -Larsson Spargelfeldstraße 191, AT-1126 Vienna, Austria *Head: Mrs. Johanna Folsner - Contact:Mrs. Johanna Folsner*
BELGIUM	Institute of Public Health, ISSP-LP Rue J. Wystman, 14, BE-1050 Bruxelles, Belgium *Head: Mr. Johan Peters - Contact: Mr. Fabien Bolle*
REPUBLIC OF CYPRUS	Laboratory for Control of Food Contact Materials and Control of Toys Ministry of Health, State General Laboratory (SGL) 44 Kimonos st, Acropolis, CY-1451 Nicosia, Cyprus *Head:. Mrs. Irini Polyniki - Contact: Mrs. Irini Polyniki*
CZECH REPUBLIC	NIPH- NRL for Food Contact Materials and for Articles for children under 3 years old, National Institute of Public Health (SZU') Šrobárova 48, CZ-100 42 Praha 10, Czech Republic *Head: Mrs. Jitka Sosnovcová - Contact: Mrs. Jitka Sosnovcová*
DENMARK	Department of Food Chemistry, National Food Institute Technical University of Denmark Mørkhøj Bygade 19, DK-2860 Søborg, Denmark *Head: Mr.. Jørn Smedsgård - Contact: Mr. Jens Højslev Petersen*
ESTONIA	Health Protection Inspectorate - Central Laboratory of Chemistry 2 Kotka street, EE-11315 Tallinn, Estonia *Head: Mr. Aare Laht - Contact: Mrs. Kristi Mägi*
FINLAND	Finnish Customs Laboratory Teknikantie 13, FIN-02150 Espoo Finland *Head: Mr. Heikki Pyysalo - Contact: Mrs.. Leena Partanen, Mrs. Aija Mertälinen*
FRANCE	Center for Energy Material and Packaging - Laboratoire National d'Essais 29, avenue Roger Hennequin, F-78197 Trappes Cedex, France *Head: Mr. Patrick Sauvegrain, - Contact : Mr. Patrick Sauvegrain,* & SCL Laboratoire de Bordeaux-Pessac 3, Avenue du Docteur Albert Schweitzer, F-33608 Pessac, France *Head: Mr. Bernard Médina - Contact: Mr. Jean-Louis Caussanel, Mr. Christian Tricard, Mrs. Monique Bertoli*
GERMANY	Bundesinstitut für Risikobewertung (BFR) (Federal Institute for Risk Assessment) Postfach 33 00 13, D - 14191 Berlin, Germany *Head: Mrs. Karla Pfaff - Contact: Mr. Oliver Kappenstein*
GREECE	General Chemical State Laboratory, D' Chemical Service of Athens, Section, Laboratory of Articles and Materials in Contact with Foodstuffs 16, An. Tsocha st, GR-115 21 Athens, Greece *Head: Mrs. Evgenia Lampi - Contact: Mrs. Evgenia Lampi*
HUNGARY	National Institute of Food Hygiene and Nutrition – Dept of Food additives and contaminants, Section Food Additives and Contact Materials Gyáli út 3/A, HU-1097 Budapest, Hungary *Head:. - Contact: Mrs. Monika Csermely*
IRELAND	Public Analyst Laboratory - Sir Patrick Duns Hospital Lower Grand Canal Street, IRL-Dublin 2, Ireland *Head: Mr. Michael O'Sullivan - Contact: Mr. John Keegan*
ITALY	Istituto Superiore di Sanità, Laboratorio Esposizione e rischio da materiali, c/o Dipartimento ambiente e connessa prevenzione primaria Viale Regina Elena, 299, I-00161 Roma, Italy *Head: Mrs. Maria Rosaria Milana - Contact: Mrs. Maria Rosaria Milana*
LATVIA	National Diagnostic Centre, Laboratory of Food and Environmental Investigations (LFEI) Lejupes Str. 3, LV – 1012 Riga, Latvia *Head: Mr. Raffael Joffe - Contact: Mr. Vadims Bartkevics*
LITHUANIA	National Public Health Investigation Centre, Laboratory of Chemistry Zolyno 36, LT-10210 Vilnius, Lithuania *Head: Mr. Romualdas Brusokas - Contact: Mr. Gintautas Svilpa*
LUXEMBOURG	Laboratoire National de Santé', Division du Controle des denrées alimentaires L-2011 Luxembourg *Head: Mr. Gilbert Morris - Contact: Mrs. Carole Dauberschmidt*
MALTA	Central Science Laboratory Sand Hutton York YO41 1LZ, United Kingdom *Head: Mr. John Gilbert - Contact: Mr. Albert Gambin*
POLAND	Laboratory of Department of Food and Consumer Articles Research , National Institute of Hygiene, Chocimska 24, PL-00 791 Warsaw, Poland *Head:Mr.. Kazimierz Karlowski - Contact: Mrs. Kazimiera Cwiek-Ludwicka*
PORTUGAL	ESB-GE (Portuguese Catholic University - Biotechnology College – Packaging Department) Rua Dr. Antonio Bernardino de Almeida, PT-4200-072 PORTO, Portugal *Head: Mrs. Maria de Fatima Poças - Contact: Mrs. Maria de Fatima Poças*
SLOVENIA	National Institute of Public Health of Republic of Slovenia , Dept of Sanitary Chemistry, Grablovičeva 44, SI-1000 Ljubljana, Slovenia *Head: Mrs. Ada Hočevar - Contact: Mrs. Viviana Golja*
SLOVAK REPUBLIC	National Reference Centre and Laboratory for material and articles intended to come into contact with food, Regional Public Health Authority in Poprad (RUVZ) Zdravotnicka 3, SK-058 97 Poprad, Slovak Republic *Head: Mr. Rastislav Rosipal - Contact: Mrs. Milada Syčova'*
SPAIN	Centro Nacional de Alimentación, Agencia Espanola de Seguridad Alimentaria y Nutrición (AESAN) Carretera a Pozuelo- km 5.1, ES-28220 Majadahonda-(Madrid), Spain *Head: Mr. Fernando Tovar - Contact: Mr. Jose J. Sanchez, Mrs. Juana Bustos*
SWEDEN	National Food Administration, Food Standards Division P.O. Box 622, SE-75126 Uppsala, Sweden *Contact: Mr. Hakan Johnsson*
THE NETHERLANDS	Food and Consumer Product Safety Authority (VWA), Inspectorate for Health Protection region North Paterswoldseweg 1, 9726 BA Groningen, NL-9700 AL Groningen, The Netherlands *Head: Mr. F Dannen - Contact: Mr. Durk J. Schakel*
UNITED KINGDOM	Central Science Laboratory Sand Hutton, York YO41 1LZ, United Kingdom *Head: Mr. Martin Rose - Contact: Mrs. Emma Bradley*

European Commission

EUR 24105 EN – Joint Research Centre – Institute for Health and Consumer Protection
Title: Guidelines for performance criteria and validation procedures of analytical methods used in controls of food contact materials

A CRL-NRL-FCM Publication, 1st edition 2009
Author(s): Stefanka Bratinova, Barbara Raffael, Catherine Simoneau
Luxembourg: Office for Official Publications of the European Communities
2009 – 73 pp. – 21 x 29 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-14483-7

Abstract

Test methods for materials and articles in contact with foodstuffs are required to determine the concentration of residues of monomers in the materials themselves or to determine the concentration of individual or groups of substances in food (or food simulants) which have migrated from the food contact materials.

The Community Reference Laboratory and National Reference Laboratories for food contact materials (FCM) prepared the present Guidelines to illustrate the required performance criteria for the analytical methods applied in the laboratories for FCM and provide procedures for method validation in order to estimate their performance characteristics. The scope of these guidelines is to provide rules for the performance of the analytical methods to be used in the verification of compliance with the migration limits defined in Directive 2002/72/EC,as amended, and in accordance with Directive 82/711/EEC, as amended, and others defined in the European legislation, in order to ensure the quality and comparability of the analytical results.

The document presents 4 approaches, according to the different purpose of performance assessment.

These guidelines are intended as a dynamic document and they will evolve and expand into further editions. This is the first edition. These guidelines have been endorsed by the European Union official Network of National Reference Laboratories and approved by the EU Commission competent service DG SANCO.

This work also highlights an important deliverable for the Network of NRLs. In particular, the members of the task force “Method Performance” that have dedicated time and effort to provide input into the development of these guidelines. They are gratefully acknowledged here for their contribution: NRL-BE (Fabien Bolle, Tina n’Goy), NRL-DE (Oliver Kappenstein), NRL-DK (Jens Petersen), NRL-ES (Juana Bustos), NRL-FR1 (Patrick Sauvegrain), NRL-EL (Timokleia Togkalidou), NRL-IT (Maria Rosaria Milana), NRL-NL (Durk Schakel, Dita Kalsbeek-van Wijk), NRL-PL (Kazimiera Cwiek-Ludwicka), NRL-SI (Viviana Golja), NRL-UK (Emma Bradley). Special thanks are extended to Emma Bradley for her contribution to the editing of the document.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-24105-EN-C

JRC
EUROPEAN COMMISSION

Publications Office
Publications.europa.eu

ISBN 978-92-79-14483-7