codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00100 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 4 a) **CX/MAS 02/4**
November 2002

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Twenty-fourth Session
Budapest, Hungary, 18-22 November 2002

PROPOSED DRAFT GUIDELINES FOR EVALUATING ACCEPTABLE METHODS OF ANALYSIS

BACKGROUND

The Codex Committee on Methods of Analysis and Sampling is responsible, amongst other things, for developing specific methods of analysis and endorsing those which have been submitted by various Codex Committees. It has developed General Principles for methods of analysis which have been included in the Codex Procedural Manual. It is also recommending that a criteria approach be developed for methods of analysis included in Codex Standards and, in association with that, is developing Working Instructions on the Implementation of the Criteria Approach for Codex Committees (see paper CX/MAS 02/5).

At the Twenty-Third Session of CCMAS it was agreed that the Guidelines would be further developed to make them generally applicable to governments (see para 34 of ALINORM 01/23).

This paper is gives the first draft of such Guidelines, for discussion at the Twenty-fourth Session of CCMAS; they have been prepared in a form which will eventually be suitable for incorporation in the methods of analysis and sampling section of the Codex as Codex Guidelines.

POSSIBLE APPROACHES TO EVALUATING ACCEPTABLE METHODS OF ANALYSIS

The introduction of the criteria approach does mean that thought now has to be given to developing defining and quantifying the specific criteria required in each instance. This is often complex and alternative approach is discussed in this paper.

Thus two possible approaches to evaluating acceptable methods of analysis are:

- To identify specific performance parameters and assign numeric values to these (the traditional approach)
- To identify a “fitness-for-purpose” approach, taking all values into account by defining a single parameter – a fitness function.

A. Fitness-for-Purpose Approach

This approach is discussed in Appendix I of this paper. It may be incorporated into the Guidelines depending on the outcome of discussions at the 24th Session of CCMAS.

B. Traditional Approach

The traditional, criterion-by-criterion approach is incorporated into the Guidelines given in the Appendix II of this paper.

RECOMMENDATION

It is recommended that the alternative approaches be discussed, together with the outline draft guidelines incorporating the traditional approach to evaluating methods of analysis, at the Twenty-fourth Session of CCMAS. If there is sufficient consensus, then the approaches should be further developed and then sent to governments for comment.

APPENDIX I: FITNESS-FOR-PURPOSE APPROACH TO EVALUATING METHODS OF ANALYSIS

INTRODUCTION

Increasingly the favoured approach to ensuring the use of appropriate methods of analysis for the enforcement of Codex specifications is to prescribe a set of performance characteristics rather than methods themselves. This 'criteria-based' approach enables laboratories to use any convenient analytical method that satisfies the criteria. There are two rather different aspects of selecting a method on the basis of a list of criteria, namely (a) selecting an off-the-shelf method that, on the basis of the information available, will probably deliver acceptable results, and (b) demonstrating that the chosen method is, in fact, capable of delivering fit-for-purpose analytical results in the users laboratory.

In principle it may be argued that the only information that is required in either case is a simple 'fitness function' comprising: (a) a statement defining the range of matrix types to which the method is to be applied; (b) an algebraic expression $u = f(c)$ describing the relationship between the uncertainty of measurement and the concentration of the analyte; and (c) the concentration range over which the function in (b) is applicable. If the concentration range of interest is quite narrow, as is often the case for highly controlled industrial production, item (b) is specified by a single number. Fitness functions can describe equally well the actual performance of a specific method (what has been called the 'characteristic function' [ref]) and the uncertainty that is fit for purpose for a specific field of application (the 'fitness function'). Selection of a method comprises the comparison of the characteristic function with the fitness function, which could be carried out graphically.

FITNESS FUNCTIONS

Before selecting a method then, the laboratory first has to quantify the uncertainty that defines fitness for purpose. That is the uncertainty that minimises a loss function, balancing cost of analysis against potential losses due to incorrect decisions. While a formal decision-theory approach to that is possible [refs], an uncertainty function could simply be agreed between the laboratory and the customer on the basis of professional judgement, or might be defined by an agency representing a whole application sector of an area of Codex. As a simple example of a fitness function, $u_f = 0.05c$ specifies that the standard uncertainty should be 5% of the concentration. It is emphasised that the judgement approach should be done without reference to the capabilities of actual methods. Once this 'fitness function' has been defined, it can be used to judge whether the characteristic functions of particular documented methods are suitable. Subject to validation, a method is suitable for the application if offers to provide an uncertainty that is lower than or equal to the fitness function over its whole defined scope.

CONSTRUCTING CHARACTERISTIC FUNCTIONS FROM TRADITIONAL INFORMATION

While it is straightforward to define complete fitness functions, off-the-shelf methods are as yet seldom described by adequate characteristic functions. Hopefully that situation will change, but in the mean time it is necessary to use the fragmentary and sometimes incomplete information provided by validation under a number of traditional headings, namely:

- Accuracy
- Applicability
- Detection limit and limit of determination
- Linearity
- Precision
- Recovery
- Selectivity
- Sensitivity

These characteristics are those which have been identified by Codex as being of importance.

However, consideration should be given to integrating such information as is provided under these headings, together with judgements covering the uncertainty contributions due to aspects where no information is available, into a coherent uncertainty function. The method discussed here is to build up an estimated characteristic function starting with a skeleton obtained from precision information.

SKELETON CHARACTERISTIC FUNCTION BASED ON PRECISION

Precision is a useful starting point in estimating the uncertainty function, because the standard deviation of reproducibility s_R accounts for a large measure, often the greater part, of the total uncertainty in a measurement. [ref] s_R is the principal parameter estimated by a collaborative trial and is therefore often immediately available. Moreover, it is available as a function of concentration, because the trial is normally carried out with at least five different concentrations of the analyte. It is possible to reasonably estimate s_R values at intermediate concentrations by interpolation. When s_R values are available, other types of precision-related uncertainty are not separately required, because they are subsumed into the reproducibility. The aspects of uncertainty that are *not* included in s_R (that is, method bias and matrix variability) can be estimated separately (as indicated elsewhere [crossref]) and combined with s_R in the appropriate way.

The main problem with s_R is that it is very unlikely to be estimated well by within-laboratory experiment where the method has not been subjected to a collaborative trial. [ref] This situation can be addressed by use of several surrogate estimates. Firstly, all method validations will always include simple estimates of repeatability standard deviation and/or run-to-run standard deviation. Repeatability standard deviation s_r, estimated by *inter alia* within-run replication, can be converted into an estimate of s_R by making use of the well-founded empirical observation of Horwitz that the expected value of the ratio $s_r / s_R \approx 0.5$. [ref] If run-to-run standard deviation is available, as it may well be explicitly or in the form of internal quality control charts, that information can be used additionally or alternatively. While there is no great body of experimental evidence to support it, an expected value of $s_{run} / s_R \approx 0.8$ is a reasonable presumption. Caution is required here because naive methodology used during validation can give rise to low estimates of both s_r and s_{run}. For example, for a estimate of s_r, repeat measurements must be made on separate test portions of typical materials taken through the whole analytical procedure at intervals throughout the whole duration of a routine run, preferably intercalated among normal test materials. Those precautions may have been neglected during validation.

Any estimate of s_R from lower-level precision experiments can be reinforced by reference to the Horwitz function, $s_H = 0.02c^{0.8495}$. If the characteristic function is comparable with s_H, this gives us confidence in the estimate. If the two functions differ systematically, expert judgement is required to choose between the estimates.

Single laboratory estimates of precision uncertainty are likely to be made at only one or two concentrations of the analyte. This information may have to be converted into a functional relationship over the range required. One can proceed by noting that that, at concentrations well above the detection limit, the relative standard deviation (RSD) of a method can often be regarded as a constant. So if only one RSD is available, that could be regarded as the initial characteristic function, at least over a limited concentration range. This would provide a characteristic function of the form

Eq 1 $$u_c = Ac,$$

where *A* is a constant. If two such estimates are available, an average of the two RSDs could be used, under the assumption that both concentrations are well above the detection limit. Averaging would not be valid if the assumption of constant RSD is untenable, below say 100 times the detection limit.

FACTORS SUBSUMED INTO THE PRECISION-BASED CHARACTERISTIC FUNCTION

Some users of this idea may be worried that many traditional aspects of precision-related quality may be ignored in this set up. For example, tests for linearity, evaluation uncertainty from calibration functions, systematic aspects of calibration, and sensitivity are not mentioned. But their contributions to overall uncertainty are not ignored. **Random calibration errors** contribute to repeatability (within-run) variation and run-to-run precision. **Systematic calibration errors** (for example, those caused by incorrectly prepared stock solutions) are fully represented in reproducibility variation. **Linearity** is, of course, an important aspect of an analytical method, but lack of fit brought about by ignoring non-linearity would be represented in reproducibility variation. **Sensitivity**, the gradient of the calibration function, is for most analytical methods an essentially arbitrary quantity and plays no direct part in determining the uncertainty.

TAKING ACCOUNT OF THE DETECTION LIMIT

Regardless of exactly how detection limit is conceptualised and defined, it represents the concentration levels where the net analytical signal is comparable with the magnitude of its uncertainty. Unless we are sure that we will be working well above the detection limit, we need to incorporate detection limit information into the characteristic function. That is easily accomplished. For example, taking the detection limit c_L as the concentration corresponding to a net signal of $m + 2s$ produced by a test material containing no analyte, the standard uncertainty represented in the concentration domain would simply be numerically equal to $c_L/2$ (ignoring problems associated with the definition of uncertainty at near-zero concentrations). It is reasonable to combine this base-level contribution to uncertainty with a proportional uncertainty present at higher concentrations. This gives us a skeleton characteristic function of

Eq 2 $$u_c = \sqrt{c_L^2/4 + A^2c^2}\ ,$$

a form which has been noted experimentally in many studies.[refs?] It must be born in mind that instrumental detection limits quoted are often unrealistically low and cannot be applied to real analytical systems without due consideration.

OTHER TRADITIONAL FACTORS

The remaining traditional factors, i.e., those not so far included in the skeleton characteristic function, are **accuracy**, **applicability**, **recovery** and **selectivity**. These factors are not independent, and that circumstance allows us to simplify the discussion. For example, accuracy depends on recovery and selectivity. Applicability comprises information about *inter alia* the types of matrix covered, which also bears on accuracy and has uncertainty implications.

Often there is an uncertainty contribution caused by matrix variation *within* the defined scope of the method that has not been assessed or taken into account. An allowance for this deficit may be difficult to estimate, because the uncertainty contribution is seldom estimated in current validation practice.[ref] Therefore professional judgement is called for to estimate the uncertainty contribution. If the proposed new use of the method is *outside* the defined scope, an additional uncertainty of unknown magnitude is introduced into the budget. Again, professional judgement is required to estimate that contribution. This might be difficult. However, we must remember that these judgements are for method selection purposes only at this stage: the assumed uncertainties can be verified subsequently by validation experiments.

There is no general guidance as to whether these matrix effects should be regarded as translational or rotational, i.e., additive or multiplicative. Again judgement is required for individual cases. If for example the effect was judged to produce an extra multiplicative uncertainty of relative magnitude B, the adjusted characteristic function would take the form

Eq 3 $$u_c = \sqrt{c_L^2/4 + \left(A^2 + B^2\right)c^2}\ .$$

Recovery information also has uncertainty implications.[ref] Ideally recovery factors (which are clearly measurements with uncertainties) should have associated uncertainty estimates. If these are available they should be combined into the characteristic function in the appropriate way.

However, analysts should beware of double accounting here. Some recommended methods of estimating recovery factors might *include* contributions from matrix variation, for example if a variety of CRMs were used in the estimate.

Examples of the recommended procedure

Example 1. Short concentration range, well above detection limit, no collaborative trial data available.

The fitness function

The analyte concentration is always in the range 40-50 % m/m and the customer requires a standard uncertainty of 1.5 % m/m for fitness for purpose.

The relevant validation information available

The proposed analytical method provides the following, according to validation information. (All results are % m/m.)

- Repeat analyses of a typical test material (material 1) within run gives $\bar{x} = 41.6, s_r = 0.52$.
- The detection limit is estimated at 0.02 units.
- Use of two materials for IQC implied the following statistics:
  material 2: $\bar{x} = 25.3,\ s_{run} = 0.41$
  material 3: $\bar{x} = 52.3,\ s_{run} = 0.76$
- Analysis of ten spiked test materials estimated that the recovery of the analyte at the appropriate concentration was 80±5% relative.

Building the characteristic function

- As the relevant concentration range is small it is reasonable to regard the uncertainty as invariant with concentration.
- Material 1 is within the range and gives us an estimate of $\hat{s}_R = 0.52 / 0.5 = 1.04$.
- Material 2 is out of range and therefore ignored.
- Material 3 is just over range, so it might give an estimate on the high side, but in fact gives $\hat{s}_R = 0.76 / 0.8 = 0.95$.
- The Horwitz function at a concentration of 50% m/m gives a reproducibility estimate of $s_H = 1.1$, which is consistent with the above $\hat{s}_R$ estimates and reinforces our confidence in them.
- The detection limit is far below the required range so the zero-point uncertainty would make a negligible contribution to the uncertainty and is ignored.
- We therefore use an average of the concordant results for material 1 and material 3 to give $\hat{s}_R = 1.0$ as the skeleton function.
- We build into the uncertainty an allowance for the uncertainty on the recovery factor. The uncertainty expected on the recovery-corrected result is therefore
  $u_c = 10\sqrt{\left(\frac{1.0}{10}\right)^2 + \left(\frac{5}{80}\right)^2} = 0.12$.
- As $u_c < u_f$, the method is deemed suitable.

Example 2. Extended concentration range, no collaborative trial data.

To be completed

Example 3. Extended concentration range, collaborative trial data available.

To be completed

APPENDIX II: GUIDELINES ON METHODS EVALUATING ACCEPTABLE METHODS OF ANALYSIS

SCOPE

1. These guidelines provide a framework for evaluating acceptable methods of analysis.

2. These guidelines are intended to assist countries in the application of requirements for trade in foodstuffs in order to protect the consumer and to facilitate fair trade.

3. Laboratories involved in the evaluation must comply with Codex Guidelines CAC/GL 27 on the competence of testing laboratories involved in the import and export of foods.

4. If a method of analysis has been endorsed by Codex, then preference should be given to using that procedure.

REQUIREMENTS

5. Methods should be assessed against the following criteria by laboratories involved in the import and export control of foods:

- accuracy
- applicability (matrix, concentration range and preference given to 'general' methods)
- detection/determination limits if appropriate for the determination being considered
- linearity
- precision; repeatability intra-laboratory (within laboratory), reproducibility inter-laboratory (within laboratory and between laboratories), but generated from collaborative trial data rather than measurement uncertainty considerations
- recovery
- selectivity (interference effects etc.)
- sensitivity

6. Their definition and approach to their estimation are given below.

ACCURACY

Definition

(as a concept)

The closeness of agreement between the reported result and the accepted reference value.

Note:

The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component. {ISO 3534-1} When the systematic error component must be arrived at by a process that includes random error, the random error component is increased by propagation of error considerations and is reduced by replication.

(as a statistic)

The closeness of agreement between a reported result and the accepted reference value. {ISO 3534-1}

Note:

Accuracy as a statistic applies to the single reported final test result; accuracy as a concept applies to single, replicate, or averaged value.

Estimation

Wherever possible the use of traceable reference materials should be used to determine the accuracy of the method of analysis used.

[Swedish proposal to CEN TC275 and WG10]

If certified reference materials are used during a method evaluation exercise then the mean determined value can be compared against the mean known value by calculation of the z-score.

$$z = \frac{\left(X_{found} - X_{certified}\right)}{\sqrt{\frac{s^{2}_{found}}{n_{found}} + \frac{s^{2}_{certified}}{n_{certified}}}}$$

or, if certified reference material standard deviation data are unavailable 95% confidence limit data may be used as an estimate of certified reference material standard deviation.

$$z = \frac{\left(X_{found} - X_{certified}\right)}{\sqrt{\frac{s^{2}_{found}}{n_{found}} + \left(\frac{CI}{2}\right)^{2}}}$$

A z-score within the range|z|=2 is deemed to be satisfactory.

APPLICABILITY

Definition

The analytes, matrices, and concentrations for which a method of analysis may be used satisfactorily to determine compliance with a Codex standard.

Note:

In addition to a statement of the range of capability of satisfactory performance for each factor, the statement of applicability (scope) may also include warnings as to known interference by other analytes, or inapplicability to certain matrices and situations.

Estimation

This should detail the analytes, matrices and concentrations for which the method of analysis may be used satisfactorily to determine compliance with a Codex standard. This may also include warnings as to known interference by other analytes, or inapplicability to certain matrices and situations. The Youden approach a fractional factorial approach, is commonly used to assess applicability/ruggedness.

DETECTION/DETERMINATION LIMITS

Definition: Detection Limit

The detection limit is conventionally defined as field blank + 3s, where s is the standard deviation of the field blank value signal (IUPAC definition).

However, an alternative definition which overcomes most of the objections to the above approach (i.e. the high variability at the limit of measurement can never be overcome) is to base it on the rounded value of the reproducibility relative standard deviation when it goes out of control (where 3 σ_R = 100%; σ_R = 33%, rounded to 50% because of the high variability). Such a value is directly related to the analyte and to the measurement system and is not based on the local measurement system.

Definition: Determination Limit

As for detection limit except that 6s or 10s is required rather than 3s.

However, an alternative definition that corresponds to that proposed for the detection limit is to use σ_R = 25%. This value does not differ much from that assigned to the detection limit because the upper limit of the detection limit merges indistinguishably into the lower limit of the determination limit.

Estimation

Where measurements are made at low analyte or property levels, e.g. in trace analysis, it is important to know what is the lowest concentration of the analyte or property value that can be confidently detected by the method. The importance in determining this, and the problems associated with it, arise from the fact that the probability of detection does not suddenly change from zero to unity as some threshold is crossed. The problems have been investigated statistically in some detail and a range of decision criteria proposed.

For validation purposes it is normally sufficient to provide an indication of the level at which detection becomes problematic. For this purpose the "blank + 3*s*" approach will usually suffice. Where the work is in support of regulatory or specification compliance, a more exact approach such as that described by IUPAC and various others is likely to be appropriate. It is recommended that users quote whichever convention they have used when stating a detection limit.

Detection Limit (LOD) - Quick Reference	
What to analyse	**What to calculate from the data**
a) 10 independent sample blanks measured once each. or	*Sample standard deviation* 's' of a) sample blank values, or b) fortified sample blank values
b) 10 independent sample blanks fortified at lowest acceptable concentration measured once each	Express LoD as the analyte concentration corresponding to a) mean sample blank value + 3*s* or b) 0 + 3*s*
This approach assumes that a signal more than 3s above the sample blank value could only have arisen from the blank much less than 1% of the time, and therefore is likely to have arisen from something else, such as the measurand. Approach a) is only useful where the sample blank gives a non-zero standard deviation. Getting a true sample blank can be difficult.	
c) 10 independent sample blanks fortified at lowest acceptable concentration, measured once each	*Sample standard deviation* '*s*' of the fortified sample blank values Express LoD as the analyte concentration corresponding to sample blank value +4.65*s* (derives from hypothesis testing)
The 'lowest acceptable concentration' is taken to be the lowest concentration for which an acceptable degree of uncertainty can be achieved. Assumes a normal practice of evaluating sample and blank separately and correcting for the blank by subtracting the analyte concentration corresponding to the blank signal from the concentration corresponding to the sample signal. If measurements are made under repeatability conditions, this also gives a measure of the repeatability precision (Annex A, A20)	

The determination limit (LoQ) is strictly the lowest concentration of analyte that can be determined with an acceptable level of repeatability precision and trueness. It is also defined by various conventions to be the analyte concentration corresponding to the sample blank value plus 6 or 10 standard deviations of the blank mean.

Note: Neither LoD nor LoQ represent levels at which quantitation is impossible. It is simply that the size of the associated uncertainties approach comparability with the actual result in the region of the LoD.

Determination Limit (LoQ) – Quick Reference	
What to analyse	**What to calculate from the data**
a) 10 independent sample blanks measured once each. Getting a true sample blank can be difficult.	*Sample standard deviation 's'* of sample blank value. Express LoQ as the analyte concentration corresponding to the sample blank value plus either: i) 6*s*, or ii) 10*s*
b) Fortify aliquots of a sample blank at various analyte concentrations close to the LoD. Measure, once each, 10 independent replicates at each concentration level.	Calculate the standard deviation '*s*' of the analyte value at each concentration. Plot *s* against concentration and put assign a value to the LoQ by inspection. Express LoQ as the lowest analyte concentration which can be determined with an acceptable level of uncertainty.
Normally LoQ forms part of the study to determine working range. It should not be determined by extrapolation below the lowest concentration fortified blank. If measurements are made under repeatability conditions, a measure of the repeatability precision at this concentration is also obtained.	

LINEARITY

Definition

The ability of a method of analysis, within a certain range, to provide an instrumental response or results proportional to the quality of analyte to be determined in the laboratory sample. This proportionality is expressed by an a priori defined mathematical expression. The linearity limits are the experimental limits of concentrations between which a linear calibration model can be applied with a known confidence level (generally taken to be equal to 1%)."

Estimation

For any quantitative method, it is necessary to determine the range of analyte concentrations or property values over which the method may be applied. Note this refers to the range of concentrations or property values in the solutions actually measured rather than in the original samples. At the lower end of the concentration range the limiting factors are the values of the limits of detection and/or quantitation. At the upper end of the concentration range limitations will be imposed by various effects depending on the instrument response system.

Within the working range there may exist a linear response range. Within the linear range signal response will have a linear relationship to analyte concentration or property value. The extent of this range may be established during the evaluation of the working range. Note that regression calculations on their own are insufficient to establish linearity. To do this a visual inspection of the line and residuals may be sufficient; objective tests, such as 'goodness-of-fit' tests, are better still. In general linearity checks require points at at least 10 different concentrations/property values.

Evaluation of the working and linear ranges will also be useful for planning what degree of calibration is required when using the method on a day-to-day basis. It is advisable to investigate the variance across the working range Within the linear range, one calibration point may be sufficient, to establish the slope of the calibration line. Elsewhere in the working range, multi-point (preferably 6+) calibration will be necessary. The relationship of instrument response to concentration does not have to be perfectly linear for a method to be effective but the curve should be repeatable from day to day. Note that the working and linear range may be different for different matrices according to the effect of interferences arising from the matrix.

Working and Linear Range - Quick Reference			
Analyse	***Repeats***	**What to calculate from the data**	**Comments**
1. Blank plus reference materials or fortified sample blanks at various concentrations	1	Plot measurement response (y axis) against measurand concentration (x axis). Visually examine to identify approximate linear range and upper and lower boundaries of the working range.	Ideally the different concentrations should be prepared independently, and not from aliquots of the same master solution.
Need at least 6 concentrations plus blank		Then go to 2.	This will give visual confirmation of whether or not the working range is linear. This stage is necessary to test a working range, thought to be linear and where it is intended to use single point calibration.
2. Reference materials or fortified sample blanks at at least 6 different concentrations within the linear range	3	Plot measurement response (y axis against measurand concentration (x axis). Visually examine for outliers that may not be reflected in the regression. Calculate appropriate regression coefficient. Calculate and plot residual values (difference between actual y value and the y value predicted by the straight line, for each x value). Random distribution about the straight line confirms linearity. Systematic trends indicate non-linearity.	It is unsafe to remove outliers without first checking using further determinations at nearby concentrations. If variance of replicates is proportional to concentration then use a weighted regression calculation rather than a non-weighted regression. In certain circumstances it may be better to try to fit a non-linear curve to the data. Functions higher than quadratic are generally not advised.
		Then go to 3.	
3. As for LoQ (b)		As for LoQ. LoQ effectively forms the lower end of the working range.	Work with successively lower concentrations until the accuracy and precision becomes unacceptable.

PRECISION CHARACTERISTICS

Definitions

The closeness of agreement between independent test results obtained under stipulated conditions {ISO 3534-1}

Notes: {ISO 3534-1}

1. Precision depends only on the distribution of random errors and does not relate to the true value or to the specified value.

2. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

3. "Independent test results" means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.

Repeatability [Reproducibility]: Precision under repeatability [reproducibility] conditions. {ISO 3534-1}

Repeatability conditions: Conditions where test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. {ISO 3534-1}

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. {ISO 3534-1}

Note:

When different methods give test results that do not differ significantly, or when different methods are permitted by the design of the experiment, as in a proficiency study or a material-certification study for the establishment of a consensus value of a reference material, the term "reproducibility" may be applied to the resulting parameters. The conditions must be explicitly stated.

Repeatability [Reproducibility] standard deviation: The standard deviation of test results obtained under repeatability [reproducibility] conditions. {ISO 3534-1}

Notes: {ISO 3534-1}

1. It is a measure of the dispersion of the distribution of test results under repeatability [reproducibility] conditions.

2. Similarly "repeatability [reproducibility] variance" and "repeatability [reproducibility] coefficient of variation" could be defined and used as measures of the dispersion of test results under repeatability [reproducibility] conditions.

Repeatability [Reproducibility] limit: The value less than or equal to which the absolute difference between two test results obtained under repeatability [reproducibility] conditions may be expected to be with a probability of 95%. {ISO 3534-1}

Notes:

1. The symbol used is r [R]. {ISO 3534-1}

2. When examining two single test results obtained under repeatability [reproducibility] conditions, the comparison should be made with the repeatability [reproducibility] limit

$$r\,[R] = 2.8\ s_r[s_R].$$ {ISO 5725-6, 4.1.4}

3 When groups of measurements are used as the basis for the calculation of the repeatability [reproducibility] limits (now called the critical difference), more complicated formulae are required that are given in ISO 5725-6:1994, 4.2.1 and 4.2.2.

Estimation

The calculated repeatability and reproducibility values can be compared with existing methods and a comparison made. If these are satisfactory then the method can used as a validated method. If there is no method with which to compare the precision parameters then theoretical repeatability and reproducibility values can be calculated from the Horwitz equation for concentrations down to 120 µg/kg or the modified equation at levels less than 120 µg/kg and greater than 13.8%.

i.e.

$$s = 0.22c \qquad if \;\; c < 1.2\times10^{-7}$$
$$s = 0.02c^{0.8495} \qquad if \;\; 1.2\times10^{-7} \le c \le 0.138$$
$$s = 0.01c^{0.5} \qquad if \;\; c > 0.138$$

Definition

Proportion of the amount of analyte present or added to the test material which is extracted and presented for measurement.

Estimation

Analytical methods do not always measure all of the analyte of interest present in the sample. Analytes may be present in a variety of forms in samples not all of interest to the analyst. The method may thus be deliberately designed to determine only a particular form of the analyte. However a failure to determine all of the analyte present may reflect an inherent problem in the method. Either way, it is necessary to assess the efficiency of the method in detecting all of the analyte present.

Because it is not usually known how much of a particular analyte is present in a test portion it is difficult to be certain how successful the method has been at extracting it from the matrix. One way to determine the efficiency of extraction is to spike test portions with the analyte at various concentrations, then extract the fortified test portions and measure the analyte concentration. The inherent problem with this is that analyte introduced in such a way will probably not be held as strongly as that which is naturally present in the test portion matrix and so the technique will give an unrealistically high impression of the extraction efficiency. It is however the most common way of determining recovery efficiency, and it is recognised as an acceptable way of doing so. However the drawback of the technique should be borne in mind. Alternatively it may be possible to carry out recovery studies on reference materials, if suitable materials are available. Provided these have been produced by characterisation of natural materials rather than by characterisation of synthetic materials into which the analyte has been spiked, then the recovery study should accurately represent the extraction of real test portions.

Recoveries - Quick Reference

Analyse	Repeats	What to calculate from the data	Comments
Matrix blanks or samples unfortified and fortified with the analyte of interest at a range of concentrations	6	Determine recovery of analyte at the various concentrations. Recovery (%) = (C1-C2)/C3 X 100 Where, C1 = concentration determined in fortified sample C2 = concentration determined in unfortified sample C3 = concentration of fortification	Fortified samples should be compared with the same sample unfortified to assess the net recovery of the fortification. Recoveries from fortified samples or matrix blanks will usually be better than real samples in which the analyte is more closely bound.
Certified reference materials (CRM)		Determine recovery of analyte relative to the certified value	Depending on how the CRM was produced and characterised, it may be possible to get >100% recovery.

SELECTIVITY

Definition

Selectivity is the extent to which a method can determine particular analyte(s) in mixtures or matrices without interferences from other components.

Selectivity is the recommended term in analytical chemistry to express the extent to which a particular method can determine analyte(s) in the presence of interferences from other components. Selectivity can be graded. The use of the term specificity for the same concept is to be discouraged as this often leads to confusion.

Estimation

Selectivity/specificity are measures that assess the reliability of measurements in the presence of interferences. The selectivity of a method is usually investigated by studying its ability to measure the analyte of interest in test portions to which specific interferences have been deliberately introduced (those thought likely to be present in samples). Where it's unclear whether or not interferences are already present, the selectivity of the method cab be investigated by studying its ability to measure the analyte compared to other independent methods/techniques.

Confirmation of identity and selectivity/specificity - Quick Reference

What you do	How many times	Calculate / determine	Comments
Analyse samples, and reference materials by candidate and other independent methods.	1	Use the results from the confirmatory techniques to assess the ability of the method to confirm analyte identity and its ability to measure the analyte in isolation from other interferences.	Decide how much supporting evidence is reasonably required to give sufficient reliability.
Analyse samples containing various suspected interferences in the presence of the analytes of interest.	1	Examine effect of interferences – does the presence of the interferent enhance or inhibit detection or quantification of the measurands.	If detection or quantitation is inhibited by the interferences, further method development will be required.

SENSITIVITY

Definition

Change in the response divided by the corresponding change in the concentration of a standard (calibration) curve; i.e., the slope, s_i, of the analytical calibration curve.

Note:

This term has been used for several other analytical applications, often referring to capability of detection, to the concentration giving 1% absorption in atomic absorption spectroscopy, and to ratio of found positives to known, true positives in immunological and microbiological tests. Such applications to analytical chemistry should be discouraged.

Notes: {IUPAC-1987}

1. A method is said to be sensitive if a small change in concentration, c, or quantity, q, causes a large change in the measure, x; that is, when the derivative dx/dc or dx/dq is large.

2. Although the signal s_i may vary with the magnitude of c_i or q_i, the slope, s_i, is usually constant over a reasonable range of concentrations. s_i may also be a function of the c or q of other analytes present in the sample.

Estimation

This is effectively the gradient of the response curve, i.e. the change in instrument response that corresponds to a change in analyte concentration. Where the response has been established as being linear with respect to concentration, i.e. within the linear range of the method, and the intercept of the response curve has been determined, sensitivity is a useful parameter to calculate and use in formulae for quantitation. Sensitivity is sometimes used to refer to limit of detection but this use is not generally recommended.

[Note: much of the detailed recommendations in Appendix II have been taken from published texts, specifically:

AOAC-I Peer Verified Methods, Policies and procedures, 1993, AOAC International, 2200 Wilson Blvd., Suite 400, Arlington, Virginia 22201-3301, USA.

W. J. Youden; Steiner, E. H. 'Statistical Manual of the AOAC-Association of Official Analytical Chemists', AOAC-I, Washington DC, 1975, p35.

"The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics" Eurachem Guide, 1998, http://www.eurachem.ul.pt/guides/valid.pdf.

Nomenclature in evaluation of analytical methods, including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure & Appl. Chem.*, 1995, **67**, 1699-1723.

Detection in Analytical Chemistry – Importance, Theory and Practice. L. A. Curries, ACS Symposium Series 361, American Chemical Society, Washington DC 1988. Various chapters are recommended, particularly Ch4 (Kirchmer, C. J.) and Ch 16 (Kurtz, D. A. *et al*.)

Analytical Methods Committee, "Recommendation for the Definition, Estimation and Use of the Detection Limit", *The Analyst*, 1987, **112**, 199-204.

"Evaluation of Analytical Methods used for Regulation of Foods and Drugs", W. Horwitz, *Anal. Chem.* 1982, 54 (1), 67A - 76A.

M. Thompson, *Analyst*, 2000, 125, 385-386.]