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 6b) **CX/MAS 04/7**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Twenty-fifth Session
Budapest, Hungary, 8-12 March 2004

CRITERIA FOR EVALUATING ACCEPTABLE METHODS OF ANALYSIS

CONSIDERATION OF THE FITNESS-FOR-PURPOSE APPROACH TO EVALUATING METHODS OF ANALYSIS

(Prepared by the United Kingdom)

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).

It has discussed, at the Twenty-fourth Session of CCMAS, two possible approaches to evaluating acceptable methods of analysis.

The 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.

Proposed draft Guidelines based on the former approach have been circulated for comment at Step 3 whilst it was agreed that the Delegation of the United Kingdom, with the assistance of a drafting group (Austria, Finland, France, Japan, Netherlands, Switzerland, United States) would revise the "fitness for purpose" approach discussed at the Twenty-forth Session for further consideration at the next session.

This approach is further elaborated in Appendix I.

RECOMMENDATION

It is recommended that the fitness-for-purpose approach be discussed at the Twenty-fifth Session of CCMAS. If there is sufficient consensus, then the approach should be further developed and then sent to governments for comment.

APPENDIX I: GUIDELINES FOR EVALUATING ACCEPTABLE METHODS OF ANALYSIS USING A FITNESS-FOR-PURPOSE APPROACH

[Definitions of terms used in these Guidelines

Uncertainty function: algebraic relationship describing how uncertainty of measurement varies with the concentration of the analyte in the context of an actual or hypothetical analytical procedure applied to a defined class of test material.

Fitness function: uncertainty function that specifies levels of uncertainty regarded as fit for purpose.

Characteristic function: uncertainty function that describes the performance of a defined analytical procedure applied to a defined class of test material.]

INTRODUCTION

The currently favoured approach to ensuring the use of appropriate methods of analysis is to specify 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 (fit-for-purpose) results; and

(b) demonstrating that the chosen method is, in fact, capable of delivering fit-for-purpose analytical results in the user's laboratory.

The former is the main theme of these Guidelines. The latter comprises the area of method validation, which has been discussed in parallel Codex Guidelines (see ALINORM 03/23, Appendix VII).

In principle, the only information that is required is a simple 'uncertainty function' comprising:

(a) a statement defining the analyte and 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 domain (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 the output of highly-controlled industrial production, or where a Codex limit is considered, the uncertainty can be regarded as invariant and items (b) and (c) are specified by a single number.

Uncertainty 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'). In this context, the selection of a method comprises the comparison of its characteristic function with the fitness function over the range defined in (c) above, which could be carried out numerically or graphically (Fig 1).

FITNESS FUNCTIONS

According to this scheme, before selecting a method, the analyst has to quantify the uncertainty that defines fitness for purpose. That is the uncertainty that minimises a loss function that balances the cost of analysis against potential losses due to incorrect decisions. While a formal decision-theory approach to that idea is possible [refs], an uncertainty function could simply be agreed between the laboratory and the customer on the basis of experience and professional judgement, or might be defined by an agency representing a whole application sector of chemical analysis. As a simple example of a fitness function,

Eq. 1 $$u_f = 0.05c$$

specifies that the standard uncertainty u_f should be 5% of the concentration *c*. 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 it offers to provide an uncertainty that is lower than or equal to the fitness function over its whole defined scope. (If its

characteristic function provides a *considerably* lower uncertainty than the fitness function, however, it may be that the proposed method is unnecessarily accurate and therefore unnecessarily expensive.)

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 we have to make do with the fragmentary and sometimes incomplete information provided by validation under a number of traditional headings [ref], namely: Accuracy, Applicability, Detection limit and limit of determination, Linearity, Precision, Recovery, Selectivity, and Sensitivity. These Guidelines integrate these functions into a single coherent uncertainty function such information as is provided under these headings, together with judgements covering the uncertainty contributions due to aspects where no information is available. The method advocated 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 σ_R accounts for a large measure, often the greater part, of the total uncertainty in a measurement. [ref] σ_R is the principal parameter estimated by a collaborative trial (interlaboratory method performance study) and is therefore often immediately available. Moreover, it is available as a function of concentration, because the collaborative trial is normally carried out with at least five different materials containing a range of concentrations of the analyte. It is possible to reasonably estimate σ_R values at intermediate concentrations by interpolation. When σ_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 σ_R (that is, method bias and matrix variability) can be estimated separately (as indicated elsewhere [cross ref]) and combined with σ_R in the appropriate way.

The main problem with σ_R is that it is very unlikely to be estimated well by within-laboratory experiments in cases where the method has not been subjected to a collaborative trial. [ref]. This situation can be ameliorated by the use of one or more surrogate estimates of σ_R. Firstly, method validations will always include simple estimates of repeatability standard deviation and/or run-to-run standard deviation. Repeatability standard deviation σ_r, estimated by within-run replication, can be converted into an estimate of σ_R by making use of the well-founded empirical observation of Horwitz that the expected value of the ratio

Eq. 2 $\sigma_r / \sigma_R \approx 0.5$. [ref]

If run-to-run standard deviation σ_{run} is available, as it may well be explicitly or implicitly 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

Eq. 3 $\sigma_{run} / \sigma_R \approx 0.8$

is a reasonable presumption.

Caution is required here because the naive methodology often used during method validation can give rise to low estimates of both σ_r and σ_{run}. For example, for a satisfactory estimate of σ_r, repeat measurements must be made on separate test portions of typical materials taken through the complete analytical procedure at random 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 σ_R derived from lower-level precision experiments can be checked by reference to the Horwitz function,

Eq. 4 $\sigma_H = 0.02c^{0.8495}$.

If the skeleton characteristic function is comparable with σ_H, the concurrence gives us confidence in the estimate. If the two functions differ systematically, expert judgement is required to choose between the estimates but, in the absence of sound evidence to the contrary, the higher indicated level of uncertainty is likely to be more correct.

Within-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. 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, where the concentration is well above the detection limit, that RSD 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 5 $$u_c = Ac,$$

where *A* is a constant. If two or more such estimates are available, an average of the RSDs could be used, again under the assumption that all analyte concentrations are well above the detection limit. Averaging would not be valid if the assumption of constant RSD is untenable, below say 50 times the detection limit.

FACTORS SUBSUMED INTO THE PRECISION-BASED CHARACTERISTIC FUNCTION

Potential users of this idea may be worried that many traditional aspects of precision-related quality may be ignored in the foregoing set up. For example, tests for linearity, evaluation uncertainty derived from calibration data, systematic errors 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 bias due to 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. It is necessary to incorporate detection limit information into the characteristic function unless working well above the detection limit. That is easily accomplished. For example, taking the detection limit c_L as the concentration corresponding to a net signal of $\mu + 2\sigma$ 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). Combining this base-level uncertainty contribution with a putative proportional uncertainty present at higher concentrations provides a plausible comprehensive model of uncertainty, with support from substantial amount of empirical data. By using the usual rule for combining independent uncertainties, this model gives us a skeleton characteristic function of

Eq. 6 $$u_c = \sqrt{c_L^2/4 + A^2c^2},$$

a form that has been noted experimentally in several studies.[refs?]

When specifying the parameters of Eq 6, detection limits quoted in the literature are usually 'instrumental detection limits', *i.e.*, they represent only instrumental precision under the best possible conditions and exclude any other, often much more substantial, sources of error. They are therefore unrealistically low and cannot be applied to real analytical systems without due consideration. For practical applications estimates of detection limits under reproducibility conditions should be used.

There is no necessity for the fitness function to specify a baseline uncertainty at zero concentration. However, it will often be the case that there exists a level of uncertainty, below which it is unnecessary to go, however small the analyte concentration falls. In such instances, fitness functions of the form of Eq 6 can be employed.

OTHER TRADITIONAL ASPECTS OF VALIDATION

The remaining traditional aspects, *i.e.*, those not so far included in the skeleton characteristic function, are accuracy, applicability, recovery and selectivity. These factors are not independent, a circumstance that 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.

Sometimes 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 may be called for in the estimation of the uncertainty contribution. If the proposed new use of the analytical method is *outside* the defined scope of the original validation, an additional uncertainty of unknown magnitude is introduced into the budget. Again, professional judgement is required to estimate that contribution. This might be difficult. In the present context, these judgements are for method selection purposes only: 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 matrix effect was judged to produce an extra multiplicative uncertainty of relative magnitude *B*, the adjusted characteristic function would take the form

Eq. 7 $$u_c = \sqrt{c_L^2/4 + \left(A^2 + B^2\right)c^2} \ .$$

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.

APPENDIX: EXAMPLES OF THE APPLICATION OF THE RECOMMENDED PROCEDURE.

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

Scenario

The analyte concentration is always in the range 40-50 % m/m.

The fitness function

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 and IQC 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.
- 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 95±2% 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 defined concentration range and by Eq. 2 gives us an estimate of $\hat{\sigma}_R = 0.52 / 0.5 = 1.04$.
- Material 2 is out of range and therefore should be ignored but, in fact, it provides an RSD similar to that of Material 3 (as expected when the concentration is well above the detection limit), and this adds confidence to the $\hat{\sigma}_R$ value derived from Material 3.
- Material 3 is just over range, so it might give an estimate on the high side, but in fact (by Eq. 3) gives $\hat{\sigma}_R = 0.76 / 0.8 = 0.95$.
- The Horwitz function (Eq. 4) at a concentration of 50% m/m gives a reproducibility estimate of $\sigma_H = 1.1$, which is consistent with the above $\hat{\sigma}_R$ estimates and reinforces our confidence in them.
- The detection limit is far below the required range so the zero-point uncertainty makes a negligible contribution to the uncertainty and is ignored.
- Use a consensus of the concordant results for material 1 and material 3 to give $\hat{\sigma}_R = 1.0$ as the skeleton function.
- Incorporate into the uncertainty an allowance for the uncertainty on the recovery factor. The uncertainty expected on the recovery-corrected mid-range result (45% m/m) is therefore $u_c = 45\sqrt{\left(\frac{1.0}{45}\right)^2 + \left(\frac{2}{95}\right)^2} = 1.4$.

Comparison of uncertainty functions

As $u_c < u_f$, the method is deemed suitable.

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

Scenario

A commodity is being sold on the basis that the concentration of a particular constituent falls between 5 and 50 ppm. Experience has shown that batches analysed can have levels of the contaminant over a somewhat wider range than that.

The fitness function

Preliminary considerations suggest that an RSU of about 10% over the range indicated and down to 2 ppm would meet requirements, so the fitness function is $u_f = 0.1c, \quad 2 \le c \le 50$.

The relevant validation information available about the candidate method

- The instrumental detection limit of the method gleaned from the literature was 0.25 ppm when adjusted for the dilution of the test portion after chemical treatment.
- The repeatability standard deviation, estimated from 10 individual test portions of two putatively typical test materials in a single run, was reported as follows:
  $\bar{x}_A = 31, \quad s_A = 1.4, \quad RSD = 0.045;$
  $\bar{x}_B = 103, \quad s_B = 2.6, \quad RSD = 0.033.$
- Single determinations of the analyte in five certified reference materials provided the following results.

Certified value	Uncertainty on certified value	Reported value
8.0	0.4	7.2
108	5.1	101.3
21.5	0.8	22.7
42.1	0.9	40.9
20.2	1.0	20.4

Building the characteristic function

- Realistic detection limits are often much higher than instrumental values, and we adopt the arbitrary decision to raise it by a factor of five to 1.25 ppm. As a baseline standard deviation, its contribution is 1.25/2 = 0.625, and that may not be negligible at the low end of the concentration range of interest, so it will be included in the model.
- The RSDs of the two materials are comparable, and the lower concentration is apparently about 100 times the detection limit, so it is a reasonable assumption that both are estimates of the same constant RSD, so adopt the mean of the two RSDs (namely 0.039) multiplied by 2.0 (according to Eq 2) as the likely value of *A* in Eq. 5, 6 etc. Under these assumptions, the skeleton characteristic function is given by

  $u_c = \sqrt{0.625^2 + 0.078^2 c^2}$.

  This is reasonably consistent with predictions from the Horwitz function. For example, at 10 ppm the predicted standard uncertainty is $u_c = 1.0$ and $\sigma_H = 1.1$, while at 50 ppm $u_c = 3.9$ and $\sigma_H = 4.4$ ppm. Accordingly we feel confident in using the skeleton function.
- Considering the results on the reference materials, a reasonable way of summarising them would be obtained by looking at the relative differences between the certified values and the found values, i.e., $\left(x_{found} - x_{cert}\right) \Big/ x_{cert}$ which gives the following (in the original order): -0.10; -0.06; 0.08; -0.03; 0.02. If the mean of these values were significantly different from zero (indicating evidence of overall bias) or the standard deviation were substantially greater than the relative uncertainty expected, the result would suggest that there was an additional source of uncertainty, perhaps due to matrix variation, that needed to be taken into account. In fact the mean relative deviation is zero and the standard deviation is 0.053. The expected relative standard uncertainty (RSU) is obtained by combining the mean RSU of the certified values (0.04) with the relative standard deviation of run-to-run precision, which is estimated as 0.8 × 0.078 = 0.062. Thus the expected RSU is $\sqrt{0.053^2 + 0.062^2} = 0.08$. This is greater than the observed value, so there are no compelling grounds for inflating the skeleton function. (In fact we might harbour suspicions that the results on the reference materials were unnaturally accurate.)

- This gives a final characteristic function of

  $u_c = \sqrt{0.625^2 + 0.078^2 c^2}$.

  Comparison of uncertainty functions

Fig. 1 shows the plots of u_c and u_f. Over most of the designated concentration range, the characteristic function is below the fitness function, showing that the proposed analytical method could apparently deliver the required degree of accuracy. However, at concentrations below about 10 ppm, the characteristic function is too high, as a consequence of the baseline variability. In order to meet the fitness-for-purpose requirement over the whole range, a method with a lower detection limit would be required.

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

Scenario

The requirement addresses the determination of a trace constituent that usually occurs at concentrations between 10 and 100 ppm. From general experience with similar tasks, the nature of the test material and of the proposed method of determination suggest that uncertainty due to matrix variation within the defined class would be restricted to less than 3% of the concentration. Recovery is expected to be 100%, because there is no scope for loss of analyte.

The fitness function

The client specifies that the uncertainty on the result should not exceed 10% of the concentration or 5 ppm, whichever is the greater.

The relevant validation information available

A collaborative trial has been carried out and the results, on a recovery corrected basis, are as follows.

Test material	Concentration	σ_R
A	16.0	1.2
B	31.4	2.0
C	39.8	2.5
D	42.9	3.7
E	46.6	3.8
F	57.1	3.4
G	63.2	4.4
H	69.9	4.0
I	88.6	7.1
J	94.3	5.1

Building the characteristic function

A plot of the collaborative trial data shows the trend of the σ_R as increasing with the concentration of the analyte (Fig.2). Fitting the data to Eq. 6 provides the estimates $\sigma_R(0) = 1.3$, $c_L = 2.6$ and $A = 0.064$. The resulting relationship, the skeleton characteristic function, given by $u_c = \sqrt{1.68 + 0.0041c^2}$, is shown as a bold line. (The Horwitz function (fine line) is shown for comparison, and predicts somewhat higher σ_R than observed. In this instance we can ignore the discrepancy as we have collaborative trial data.) Incorporate an extra rotational uncertainty to account for matrix variations equivalent to the maximum thought likely in this analytical system, i.e., 5% relative, which, using Eq.7 gives the final characteristic function $u_c = \sqrt{1.68 + \left(0.05^2 + 0.0041\right)c^2}$ or $u_c = \sqrt{1.68 + 0.0066c^2}$.

Comparison of uncertainty functions

The characteristic function and the fitness function are shown in Fig. 3. The characteristic function is seen to be the lower over the whole of the relevant range (10-100 ppm), so the method is apparently fit for purpose.

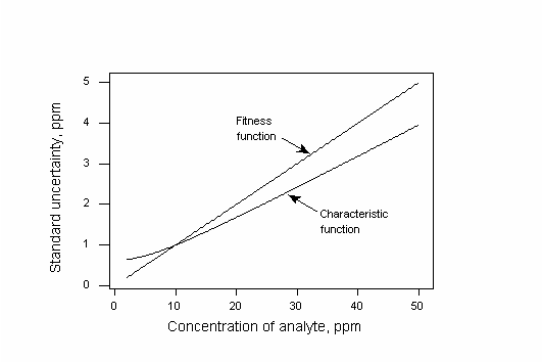

Fig 1. Fitness and characteristic functions, Example 2.

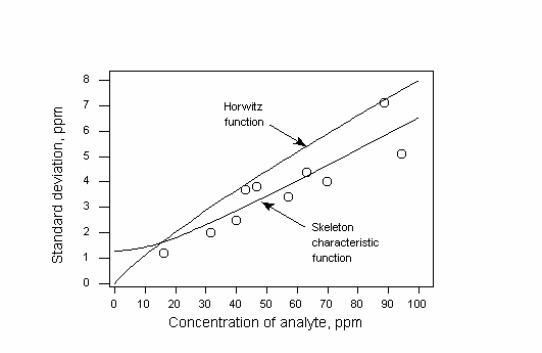

Fig.2. Collaborative trial results (circles) from Example 3, with fitted characteristic function and Horwitz function.

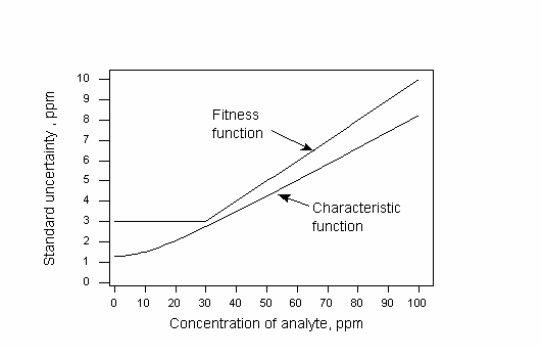

Fig.3. Fitness and characteristic functions, Example 3.