codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00153 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 3a) **CX/MAS 09/30/3**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Thirtieth Session
Balatonalmádi, Hungary, 9 - 13 March 2009

CRITERIA FOR EVALUATING ACCEPTABLE METHODS OF ANALYSIS

DRAFT GUIDELINES FOR EVALUATING ACCEPTABLE METHODS OF ANALYSIS (At Step 7)

(prepared by New Zealand)

Background

1. The Draft Guidelines for Evaluating Acceptable Methods of Analysis were advanced to Step 7 at the 26th Session of CCMAS, and since then have been considered by two working groups led by New Zealand, the results of which were considered at the 27th and 28th Sessions[1].

2. The working groups revised the core guidelines, developed new criteria for evaluation of methods, recommended further work on statistical procedures, and recommended updating Codex guidelines for method-performance studies[2].

3. At the 28th and 29th Sessions the Guidelines were held at Step 7 pending the scientific justification of the approach advocated by New Zealand, by publication in a peer-reviewed journal. At the 29th Session[3], New Zealand reported[4] that three papers had been drafted, one of which had been submitted for publication. The committee agreed to retain the draft Guidelines at Step 7 until publication of the three papers in scientific journals.

Progress on the Papers

4. The paper submitted for publication was entitled, "Allowing for imprecision in experimental estimates of measurement uncertainty." This is an introductory paper that contained technical material needed for the proposals for method evaluation which were intended to follow in two subsequent papers. The paper has been rejected, mainly because the work was regarded as "a minimal contribution compared to existing techniques". This is probably a fair comment, since the work is based on the well-established concept of tolerance intervals, only applied to the problem of evaluation of test methods. The use of tolerance intervals is well established, though it is not widely known and not used in the area of product compliance assessment.

5. New Zealand therefore concludes that there is little merit in pursuing publication of this paper, simply for the purpose of establishing credibility through peer assessment, when the techniques are already well established. We are considering how best to proceed with the remaining papers. The matter at issue is the principles on which our approach relies, rather than one of verifying the technical details of the statistical theory and calculations by which the approach

---

1 ALINORM 06/29/23, paras 12-21 and ALINORM 07/30/23, paras 18-27
2 CX/MAS 07/28/3
3 ALINORM 08/31/23, paras 16-19
4 CX/MAS 08/29/3

may be implemented. The latter are matters which, to a considerable extent, are already dealt with in statistical literature, and it is not clear that we are in a position to add significantly to this material, or indeed that there is a need to do so.

Outline of the Approach

6. We consider the primary requirement of an analytical method to be one of fitness for purpose. Methods are accepted by Codex to enable assessment of products against Codex standards. So in the contexts addressed here the "purpose" of an analytical method is to test product for compliance with specified limits on analyte levels.

7. Therefore it seems appropriate that any decision on acceptance of a method should be based on how the performance of the method affects assessments of conformity, and fitness for purpose must be judged in the context of the entire test protocol. That is, the numbers of samples and the various calculations that are to be performed to reach a decision need to be known, to enable the effect of measurement error to be assessed.

8. Such details are given in a sampling plan, of the sort given, for example, in the Codex General Guidelines on Sampling. However the sampling plans given in the General Guidelines do not currently allow for the presence of significant measurement error, and we suggest elsewhere that this is a matter requiring attention from CCMAS. We have therefore suggested some possible general principles for compliance testing in the Annex. The discussion that follows is based on these principles.

9. To satisfy these principles it is necessary to assess the effect on acceptance rates, at appropriate analyte levels, of the measurement errors associated with a particular analytical method. For example, to be sure of a high acceptance rate for compliant product it will often be necessary to allow for measurement error by setting a cut-off for measured results that is higher than the actual specification limit. If this cut-off for measured results is too high, this may result in unacceptably high acceptance rates for product with measured analyte concentrations that are well above the specification limit. Usually, the greater the imprecision associated with the analytical method, the higher the cut-off will have to be for that method, and the greater the risk to which the importing country will be exposed as a consequence. In judging whether a method is fit for purpose, attention has to be given to whether this risk is acceptable.

10. Most of the material presented to CCMAS in which measurement errors have been considered has concentrated on the effect of measurement error in judging the compliance of a single sample. This is of course a considerably less difficult problem than that of assessing compliance of a possibly heterogeneous lot, in that consideration of the sampling plan used to obtain and interpret multi-sample assessments is avoided.

11. In this simple case we suggest a method of judging the fitness for purpose of the analytical method used, to illustrate the line of approach that follows from our principles. More generally the same sort of considerations would apply, but the details would depend on the sampling plan as well as the analytical method.

12. It is noteworthy that even in this simple case, the evaluation of fitness for purpose seems to require information that is not customarily provided in an assessment of a method's performance (see the paragraph 15 below.)

Judging "fitness for purpose" in the single sample case

13. We consider a situation in which a sample is to be tested for compliance with a maximum true analyte level of ξ_0. We first determine a cut-off x_C for bias-corrected measured values, $x_C = \xi_0 + T$. The quantity *T*, which needs to be determined for the method concerned, and which plays a large part in determining the suitability of the method, is described below (paragraph 15).

14. This cut-off x_C will then determine $\xi_{0.5}$, an analyte level at which the probability of acceptance of product is approximately 50%. This level is particularly useful as the probability of acceptance when the true analyte level is equal to the cut-off is largely independent of the true precision of the method, and is therefore less uncertain. It may also be desirable to consider a further level ξ_1, the analyte level for which there is 95% confidence that the acceptance rate will be 5% or lower. Often this will be simply $x_C + T = \xi_0 + 2T$. An analytical method is then judged "fit for purpose" if it yields suitably low values for $\xi_{0.5}$ and ξ_1.

Information required in the single sample case

15. To implement the method above the following information is required:
    a. An estimate of bias *b* (except in the case of Type 1 methods). It is assumed that in accordance with Principle 3 (Annex), the measured results *X* will be adjusted to "bias-corrected" estimates $X' = X - b$. It is highly desirable that an estimate of the standard error of *b* should be available.
    b. An upper limit *T* for the measurement errors $E = X' - \xi$ (where ξ is the true analyte concentration.) There should be 95% confidence that *T* exceeds at least 95% of the analytical errors *E* generated by the method. (Principle 2.)

16. It is a serious problem that for most analytical methods, a suitable upper limit *T* may not currently be available, and unless the need for such limits is made known, method validation studies seem unlikely to routinely give such limits. Where the characteristics of the method are derived from experimental data, *T* corresponds to a one-sided tolerance limit for the measurement errors *E*, with 95% probability (or confidence) of 95% coverage. In simple cases *T* can be calculated exactly (assuming that the *E* are independently and normally distributed) from the non-central t distribution. In other cases approximate methods may be developed. In the meantime, conservative values may have to be used, bearing in mind the requirement for a confident assertion of at least 95% coverage applicable for the single sample case.

Evaluation of Methods – more general case

17. As before, in the general case the analytical method should be considered in conjunction with the sampling plan used. The points in paragraphs 18 and 19 were put forward in the Draft Guidelines as likely to be required for such a joint assessment to be carried out. When appropriate sampling plans incorporating measurement uncertainty have been investigated, it is possible some modification might be needed to this list.

18. To enable the assessment of fitness for purpose, at least the following should be reported:
    a. An estimate of method bias over the relevant range
    b. A 95% confidence interval for the method bias
    c. Estimates of the repeatability standard deviation, the between-laboratory standard deviation and the reproducibility standard deviation
    d. Upper confidence limits for the precision parameters in c). If the criteria in the Annex [to the Draft Guidelines] are to be used, the limits given should include 80% limits[5].

19. In the case when a candidate method was to be used to in place of a specified method in a compliance test, without changing the tolerances or sampling plan associated with the test, it was suggested[6] that upper 80% confidence limits for the repeatability and between-laboratories standard deviation should not exceed the accepted values of these standard deviations for the specified method by more than 14%. Automatic bias correction was also required[7]. This was based on a requirement for 80% confidence that product exposed to a 5% risk of failure under the standard method, should not be exposed to a risk of more than 7.5% under the candidate method.

---

5 CX/MAS 07/28/3, Appendix 1, para 12
6 CX/MAS 07/28/3, Annex, Criteria, Note 1
7 CX/MAS 07/28/3, Annex, Criteria, Note 2

20. The confidence level of 80% specified above is considerably less than the 95% that we have suggested in paragraph 14 above. The 80% lower level was suggested because anything higher seemed likely to yield very poor chances for a candidate method to be considered acceptable, using tolerances based on the standard method. In our view it is preferable for tolerances to be recalculated when a substitute method is used, and if this is done a higher level of confidence may reasonably be sought.

Evaluation of Methods Using the "Criteria Approach"

21. CCMAS is separately considering another set of guidelines for evaluation of methods of analysis, as part of implementing the "criteria approach" described in the Procedural Manual. Working instructions for implementation have been adopted for inclusion in the Procedural Manual, and guidelines for establishing numeric values for method criteria and for evaluating methods for compliance with the criteria are proposed[8].

22. These guidelines should be rationalized with the guidelines developed under this work item, to avoid duplication of effort. Since the guidelines under this item have been held at step 7 since the 26th session it is desirable that they should be completed without further delay.

23. Nevertheless it is also highly desirable that the criteria used for evaluation of methods should be reviewed to ensure the methods are fit for the purpose of compliance assessment. The criteria currently used carry significant risks of excluding suitable methods and accepting unsuitable methods, and the effect of any particular accepted method on producer's risk is unknown. We feel that the criteria have been formulated by considering what is scientifically reasonable, rather than by assessing how accurately the analyte level needs to be known to provide reasonable or agreed protection to both exporting and importing countries. That is, the question being asked is "how accurately can we reasonably expect to know the true analyte concentration" rather than "how accurately do we need to know the true analyte concentration." We see the latter question as more relevant to Codex. Situations where relatively 'rough and ready' methods might be appropriate and situations where no existing method gives the required protection are not catered for.

24. The current procedures simply require that an estimate of a parameter (e.g. reproducibility SD, recovery) lie within a range, with no quantifiable guarantee that the true value of the parameter does in fact lie within that range. Thus a test for product compliance based on the assumption that the parameter value lies within the range will be subject to uncontrolled error rates. To avoid this situation it is desirable that confidence limits for the true values of the precision parameters should be established.

Recommendations

25. Guidelines for evaluation of methods should be completed, taking account of the work of both Sweden, NMKL and a working group (CX/MAS 09/30/7) and the New Zealand-led working group (CX/MAS 07/28/3).

26. CCMAS should note the apparent risks of the criteria as currently formulated and the need for future work to ensure that methods of analysis are fit for purpose, including:
    a. Development of principles for compliance assessment of foods
    b. Revision of the Codex guidelines for method-performance studies
    c. Revision of the Working Instructions for the Implementation of the Criteria Approach in Codex
    d. Revision of the Guidelines.

[8] CX/MAS 09/30/7

Annex

Possible principles for compliance assessment of foods

Our treatment of evaluation of acceptable methods is based on the following possible principles.

Note: These principles are formulated merely for the purpose of illustration, in order to provide a basis for discussing the evaluation of methods of analysis. At this stage they are not a specific proposal, and such principles need separate discussion in their own right.

1. General Philosophy

a) In testing product for compliance with specification limits, it is of primary importance to control acceptance rates at various analyte levels and for various compliant and non-compliant lot compositions.
b) In general, a testing procedure should be such that lots consisting entirely of truly compliant product are subjected to a maximum probability of rejection of, say, 5%[9].
c) If a procedure meeting this requirement has been properly carried out[10] and has resulted in failure for a particular lot, then an importing county is entitled to reject that lot (subject to any applicable disputes procedure), without consideration of the extent to which the lot has or has not been proved compliant or not compliant. In other cases, the lot should be accepted.

2. Confidence required in respect of each analytical method used

It is necessary that for each individual analyte and method, there should be a reasonable level of confidence (say 95%) that the quoted minimum acceptance and coverage rates will be met or exceeded.

The rates of acceptance may depend significantly on the bias and precision characteristics of the analytical method used to measure analyte levels in samples from the lot. These will almost always be known only approximately, and could be subject to considerable errors of estimation. Adopting the view expressed in a) and b) entails that tests for product compliance should allow for the fact that the bias and imprecision of a method may have been underestimated, to an extent that is consistent with the experimental data; for example by using appropriate upper confidence limits rather than simple estimates of the method reproducibility.

3. Explicit adjustment for bias

A possible bias in the analytical result should always be allowed for (whether statistically significant or not) either by adding it to the relevant cut-off or (preferably) by subtracting it from the individual results.

---

[9] 5% seems usually to be considered as reasonable; however it could be varied according to circumstances.

[10] "properly carried out": that is, the relevant sampling plan and the protocol for the method of analysis have been adhered to.