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 8 **CX/MAS 09/30/9-Add.1**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

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

PROPOSED DRAFT REVISED GUIDELINES ON MEASUREMENT UNCERTAINTY

GOVERNMENT COMMENTS AT STEP 3

(Australia, Japan, New Zealand)

AUSTRALIA

General Comments:
Australia supports the development of guidelines on MU, particularly in regard to the interpretation of MU when assessing compliance against a limit. However, Australia strongly believes that any document developed for this purpose should not be prescriptive but rather provide information and examples of approaches that may be used.

The document developed by CCMAS should serve as an "overarching" document that provides guidance to all CODEX committees that deal with analytical results and the interpretation of their associated MUs. Other committees that may be working on this topic e.g. CCPR who are revising "Guidelines on the Estimation of Measurement Uncertainty" (CAC/GL 59-2006) - should wait until the CCMAS document is finalised so that progress of their work can be harmonised with the CCMAS document.

Specific Comments:

Section 5, second paragraph; the text in bold should be added to the end of the sentence to read: '.....approach using collaborative trial data**, in-house quality control data, validation data or a combination of these data.**' (note that this is understood from 5th bullet, bottom page 6 and last para Section 6).

Section 9.2, Third paragraph. Since recoveries will inevitably vary, this paragraph is very problematic and should be either deleted or changed to stipulate that standards should be set assuming 100% recovery, and test results be preferably corrected for recovery or reported with a measurement uncertainty enlarged to take into account any significant uncorrected bias.

Australia is concerned that sections of the "Explanatory Notes" appear to be too prescriptive or require further explanations. For example:

The example given in Section 9.1 only deals with situations where assessments are against an upper limit. However, given that a lot of situations in international trade deal with compliance within a range, additional examples of how these situations may be treated would be helpful.

In Section 10, references to export and import certificates should be removed as these terms mean different things to different countries and may not be issued or include results in all cases. Furthermore, the text should be made more generic and less prescriptive. In particular, the last sentence in the last para of this section requires further clarification to indicate that in certain circumstances, exporting only in cases where the result minus its MU is below a specification may not be possible (e.g. certain grain protectants where treatment at concentrations closer to the MRL are required to ensure adequate protection during transit). Consequently, management of compliance with an importer's specification should be left to the exporter and that the decision should be based on risk.

Australia recommends that the current sections 9 and 10 be combined and the title changed to "Consideration of MU when Comparing Results with Codex Standards' or similar. The heading to the current Section 9 is too complex. The details can be picked up in the text of a new combined Section. The new Section should cover the interpretation of MU for regulatory purposes. This is a separate issue to how MU might be used to judge whether there is a genuine dispute between test results. If it is desired to mention dispute resolution as a related issue, then this could be done via reference to the CCMAS dispute resolution guidelines.

With regard to the interpretation/consideration of MU for Codex purposes, Australia suggests the Guidelines should present the different approaches that might be taken, noting that decisions taken in the light of MU should be risk-based. Australia considers the EURACHEM/CITAC Guide on the 'Use of Uncertainty Information in Compliance Assessment', 2007 provides very useful advice on this topic and should be referenced in the body of this new Section as well as in the Useful Reference Section of the Guidelines.

JAPAN

We appreciate the United Kingdom for preparing the proposed draft revised Guidelines on Measurement Uncertainty, and we are pleased to provide the following comments for further discussion.

General Comments

Throughout the proposed draft, accreditation to the internationally recognized standard (namely, ISO/IEC 17025) is mentioned repeatedly. We would like to recall that the Codex guidelines do not require the accreditation status but the compliance with the internationally recognized standard concerning accreditation*. Therefore, the relevant sentences should be amended following that line.

> * - *Compliance with the general criteria for testing laboratories laid down in ISO/IEC Guide17025:1999 "General requirements for the competence of calibration and testing laboratories"* (CAC/GL 27)

Also, it is suggested in the proposed draft that, in estimating measurement uncertainty, consideration should be given to the whole system including sampling at a later stage. However, we just started to have a discussion on sampling uncertainty (Agenda Item 9), and we are not certain at this point if it is appropriate to consider sampling uncertainty in estimating measurement uncertainty. Moreover, in the Project Document with which we agreed at the last Session (CRD 22, the 29th CCMAS), it is clearly stated that "The Guidelines will only consider measurement uncertainty deriving from analysis and not sampling". Therefore, we believe that the aspect of sampling uncertainty should not be included in the scope of the Revised Guidelines on Measurement Uncertainty.

Specific Comments

2. Does the Measurement Uncertainty have to be Estimated in Codex?

As we stated in the General Comments, the Codex guidelines (CAC/GL 27) do not require the **accreditation status** but the **compliance** with the internationally recognized standard concerning accreditation, such as ISO/IEC 17025:2005. Also, in the "**Introduction"** part of the main text of the guidelines (CAC/GL 54), it is said that "It is important and required by ISO/IEC 17025:2005 that analysts are **aware** of the measurement uncertainty associated with each analytical result and estimates that uncertainty", and it does not state that the measurement uncertainty **must** be estimated. We believe that the description of the current Section 2 is inconsistent with or beyond the existing Codex guidelines, and thus this Section should be deleted.

3. Does the Measurement Uncertainty Apply to both Sampling and Analysis?

Again, as stated in the General Comments, it is premature to touch upon sampling uncertainty. At least at this point, only uncertainty arising from analysis should be the focus of this guideline.

Therefore, we propose to delete the current Section 3, and add the following sentence at the end of the Section 1:

> "Measurement uncertainty applies to the whole measurement process, but not to sampling process."

4. What is the relationship between Measurement Uncertainty, the Analytical Result and the Method Used to Obtain the Result?

As the content of this Section also constitutes an essential part of the Section 1. "What is Measurement Uncertainty? ", we propose to relocate the texts to the last part of the Section 1 after adding the modification to the 2nd sentence as follows;

"It is the estimation of the measurement uncertainty associated with an analytical result that is important. Measurement uncertainty is **associated with the analytical result,** not ~~associated~~ with a method **of analysis**~~, but the values that are obtained in the validation of a method may be used to estimate the uncertainty of a result in some situations~~. This differentiation between……..and competence of laboratory must be considered."

5. Procedures for Estimating Measurement Uncertainty

In the 4th paragraph, the Section 7.6.1 of the 2nd Edition of the EURACHEM is referred, and here again, sources of uncertainty that are not covered by the collaborative study data such as sampling are mentioned. As the current Guidelines (CAC/GL 54) only deal with the measurement uncertainty associated with analytical results, the procedures suitable for this scope should be referred.

6. Considerations when Estimating Measurement Uncertainty within the Context of Codex

As we stated in the General Comments, the Codex guidelines do not require the accreditation status but the compliance with the internationally recognized standard concerning accreditation. Therefore, we propose to amend the first bullet as follows:

- ~~accredited to~~ **comply with** an ~~I~~**i**nternationally recognized ~~S~~**s**tandard (now with ISO/IEC 17025 Standard); such ~~accreditation~~ **compliance** is aided by the use of internal quality control procedures,

8. Values of Measurement Uncertainty Estimation

The assessment of microbiological quality and safety in food is outside of the Terms of reference of the CCMAS (bullet point (d) of TOR). For this reason, we propose to delete the description regarding microbiological analyses (the 3rd paragraph).

10. Use of Measurement Uncertainty and Definition of a Dispute Situation

We propose to delete the entire Section as the issue is under discussion as another Agenda Item ("Draft Guidelines for Settling Disputes over Analytical (test) Results").

Editorial Comments

Section numbers are not in the correct order after the Section 6. The Section 8 should be Section 7, and so on.

NEW ZEALAND

GENERAL COMMENTS

Thank you for the opportunity to comment on the Draft Revised Guidelines on Measurement Uncertainty. We have made rather extensive comments, because in addition to the bullet points noted in the Background, we consider it is necessary that the draft:

- Should be statistically sound
- Should take account of the practicalities of compliance assessment and trade in food
- Should not impose unreasonable costs (e.g. for food manufacture or certification) or unnecessary food rejection, and
- Should recommend sampling plans that provide consumers and producers with known levels of protection.

We agree that the draft should take note of other texts developed by CCMAS and adopted by the Commission. The *General Guidelines on Sampling* include discussion of measurement error. Those Guidelines recommend that measurement errors associated with analysis should be minimised, and point out the need to consider the relative sizes of the analytical error and the sampling error when designing sampling plans. These aspects of measurement uncertainty should be covered in the Explanatory Notes.

SPECIFIC COMMENTS ON THE EXPLANATORY NOTES

Section 1

Coverage probability of the uncertainty intervals

The statement, "the range $a \pm 2u$ represents a 95% level of confidence where the true value may be found," should be clarified to avoid the possible misinterpretation that the range $a \pm 2u$ is a 95% confidence interval

for the true value. In reality, this is a tolerance interval defining the range within which 95% (nominally) of results can be said to lie. However, it needs to be clearly stated that the coverage probability[1] of the uncertainty intervals quoted for a particular method or laboratory cannot be asserted with any degree of confidence to reach or exceed 95%, and could well fall considerably short of it.

To support this view we have prepared a paper[2] which examines the coverage rates achieved by the recommended procedures. Firstly, we have examined the coverage rates actually achieved using intervals which apply coverage factors like 1.645, 2 or percentiles of Student's t-distribution to an estimate of standard deviation to obtain an uncertainty interval. Our conclusions are that the coverage obtained by such intervals has a substantial probability, for any given method of analysis, of falling considerably below the nominal level of 95%. The recommended coefficients are simply too small to achieve the nominal coverage with high probability. They achieve the desired level of coverage only on average, that is, intervals from one estimate of standard uncertainty that are too short are in some sense considered to be "balanced out" by intervals from another that are too long. We suggest a tolerance interval with 95% confidence of 95% coverage is needed, especially before proof beyond reasonable doubt can be asserted.

The reason for the substantial probability of low coverage is that experimentally determined estimates of precision are themselves subject to considerable imprecision of a statistical nature. This imprecision must be allowed for. For example, as a first approximation, an upper confidence limit for the relevant precision parameter could be used in place of its estimated value. However the suggested procedures do not allow for this imprecision: we consider this is a serious defect, that substantially reduces the usefulness of the uncertainty intervals in testing product for compliance with limits in international trade.

To confirm our interpretation of these intervals we have examined the procedures given in one of the references listed (EURACHEM/CITAC Guide Quantifying Analytical Uncertainty in Analytical Measurement). It is clear that the procedures recommended are those that we criticise: the expansion coefficients based on Student's t-distribution are applied, even when the estimate of precision is experimentally obtained with few degrees of freedom (in the example we examined, four). In this instance, we can say, with 95% confidence, that the true standard deviation could lie within the range between 0.60 and 2.87 times the sample standard deviation estimated from such a trial.

Consequently, the coverage probability of the uncertainty intervals generated in respect of a particular analyte or method is uncontrolled. Its *expected* value may be close to 95% but there is no reason to believe that this is achieved, even approximately, in any particular case.

We have two further comments on this section. Describing the value a as "the best estimate of the true value of the concentration of the measurand" is misleading, particularly if it refers to a single analytical result. It could be more clearly put as: "The result of a measurement after correction for systematic effects."

The last sentence, about "variability around the reported results", needs to be made clearer if possible.

Section 2

The reasons why an estimate of measurement uncertainty might be requested for goods moving in international trade should be indicated, in order to avoid an increased and unnecessary demand for information on measurement uncertainty on certificates issued by competent authorities and exporters.

The Guidelines or Explanatory Notes should recommend one basis for reporting measurement uncertainty, so that users know which basis has been used. The report of uncertainty should also include how those estimates of uncertainty were derived and the uncertainties of those estimates themselves. This is particularly necessary when the estimate of uncertainty is specific to a single laboratory, as suggested in section 4.

Section 3

This explanation would be clearer if it was expressed in relation to the definition of measurement uncertainty in the *Guidelines*, for example by noting that the dispersion of values attributed to the measurand includes both analytical measurement uncertainty and sampling measurement uncertainty.

Section 4

The intention of this section should be clarified. It seems intended to prepare users for the fact that sometimes laboratories may give uncertainties that do not appear consistent with estimates from method

[1] Coverage probability: The probability that the interval will include the true value.

[2] New Zealand will make this discussion paper available as a CRD.

validation studies. It may also be read as an assurance to users that in giving an estimate of uncertainty a laboratory will have taken the relevant laboratory-specific factors into account.

The first sentence can be read in at least three different ways. It needs to be clarified.

The attempt to differentiate between "result" and "validated method" seems to create rather than remove confusion. After all, a result is generated by a laboratory, a laboratory uses a method, and method validation studies typically employ outlier screening procedures to remove any observations non-conforming to the model that assumes the same method performance across all laboratories, of which the laboratories in the study are considered a sample.

We wonder whether the perceived difference between measurement uncertainties at different laboratories has arisen because the uncertainties surrounding estimates has not been considered.

The sentence "This is expected" needs further explanation, for example by stating that the uncertainties of the estimates of *u* lead to different laboratories appearing to have different measurement uncertainties.

Last sentence: Bias and matrix should not be included in measurement uncertainty. Bias is not a measure of dispersion so its inclusion is not required, and if there is a matrix effect then the method validation study is deficient.

Section 5

The Notes should include a recommendation for laboratories to update their estimates of measurement uncertainty so as to reflect their current performance.

Paragraph 3: It would be helpful to include a reference to GL 27. We note that the Guideline does not use the term "fully" validated.

The Notes should not list procedures for estimating measurement uncertainty unless there is an assessment of their validity (for example by peer review), their suitability for Codex purposes and their availability. We have noted in our comments on Section 1 that at least one of the procedures is seriously defective in respect of its treatment of the imprecision of estimates of precision characteristics. We also consider the treatment of bias is unsatisfactory, since the same procedure advises that even where a bias is statistically significant, one can nonetheless ignore it if it is not "significant" relative to the combined uncertainty.

Section 6

It would be helpful to include a reference to GL 27. We note that the Guideline does not use the word "fully".

Section 8

It is likely that the underestimation of measurement uncertainties [with respect to their true values] which is reported in Section 8 to be of concern may in part simply arise from the fact that trials, collaborative or otherwise, may have yielded considerable underestimates of the relevant standard deviations, particularly those derived using small trials. Allowance should be made to cover this possibility, for example, by the use of an appropriate upper confidence limit for the true standard deviation.

We are concerned that these potentially uncertain estimates might be regarded as "universal", and be applied inappropriately to future assessments of similar products.

Section 9.1

The source of the quote "It is stated …" should be referenced.

The first paragraph would be improved by stating the three parties involved and what they are responsible for. The three parties are the Codex Commodity Committee, the laboratory, and the risk manager or compliance manager. The latter two are confused by the use of the term "enforcement laboratory". It seems to us that a laboratory should report only the result, together with the uncertainty, to the relevant authority in all cases. As written, the Notes can be read as advising that laboratories subtract uncertainties in some cases and not others, leading to confusion and to occasional double safety margins being used. A result that is "censored" or manipulated may give unclear direction to risk managers.

The first and last paragraphs may be read as a suggestion that Commodity Committees consider reducing the specification level to at least partly negate the effect of any allowance for measurement uncertainty. We doubt that this is appropriate: it may lead to a situation where any allowance for measurement uncertainty is purely cosmetic, the specification limit having been reduced so that the same lots are accepted and rejected as if the product were tested against the old limit with no allowance for measurement error. This would

make consideration of measurement error superfluous, except to the manufacturer, who has to work to a specification limit determined by the uncertainty of whatever measurement process the Commodity Committee has assumed the importer to use. Note moreover that it is stated that measurement error "may" not be allowed for in testing for food pathogens. There seems to be a risk that Commodity Committees will set a specification to negate the effect of a supposed allowance for measurement uncertainty that will not in fact be made.

The compliance assessment procedure described in this section is one particular way of making allowance for measurement uncertainty. It should be noted that there are other ways of allowing for measurement uncertainty: when analytical error is negligible in relation to sampling error, then sampling plans described in the *General Guidelines on Sampling* are recommended; when analytical error is larger than one third of the sampling error, or when both analytical error and sampling error are significant, there is currently no specific recommendation in Codex. Various possibilities are under consideration, including the one discussed in the proposed Explanatory Notes. It should be noted that the choice of sampling plan, taking account of the various risks, is a responsibility of the Commodity Committee.

Similarly the enforcement procedure described here is one particular way of enforcing specifications. Again it should be noted that the enforcement procedure is a responsibility of the Commodity Committee. Note that enforcement may well be carried out in the exporting country, by a certifying body.

The theoretical basis of the recommended procedures should be reviewed. We note that the international definition of measurement uncertainty (p3) seems to be based on the relative plausibility of various analyte levels given a single (?) test result, rather than the distribution of test results about a given true analyte level. The relevance of the resulting theory to one of the main problems of compliance testing, that of controlling the rates of rejection for product at various acceptable and unacceptable analyte levels, may therefore be rather questionable.

We consider that in light of the true nature of the measurement uncertainty intervals (see our comments on Section 1), Commodity Committees should not be recommended to use them in testing for compliance of imported product.

The explanatory notes should be reasonably specific about the implications of the compliance assessment procedure described in Section 9.1, including operating characteristics, for the benefit of Commodity Committees. They seem to include a) for product to have a 2.5% probability of rejection or less, it will have to have a true analyte level more than $2U$ (i.e. $4u$) below the specified limit, and b) that product with true analyte level equal to the specified limit will have a 2.5% chance of acceptance. Further, the producer's and consumer's risks will be unacceptably large in many instances. For example if we had a perfect method for which $u = 0$ (without inclusion of sampling uncertainty) there would be a 50% chance of accepting product which was 50% out of specification if a single-result assessment were used.

It also needs to be made clear that where more than one sample is taken from a lot, the approach is not applicable on a sample by sample basis. For example, an importing country may be misled into taking 10 samples from a lot on the grounds that its assessment would be more stringent. Finding one measured result for which the measured value exceeds the limit by more than the expanded uncertainty, they may then assert, as instructed by section 9.1, situation 1, that that sample is not compliant with the specification. In fact, even assuming that the uncertainty interval does have a 95% coverage rate (which we do not consider it does) the probability of finding one or more of ten measured values exceeding the limit by more than the uncertainty could be as high as 22% even when the lot is entirely compliant. It is clear that by adopting this approach an importing country could expose compliant product to an unacceptable risk of failure.

In the diagram, it is not clear what the "upper control limit" is, or what the second horizontal line is.

In Situations I to IV, it is not clear whether the maximum level (or maximum value) is the same as the maximum limit or the upper control limit.

In Situations II and IV, the sample test result is stated as giving a level of acceptance of compliance or non-compliance "beyond reasonable doubt" or "without any hesitation". The words conventionally used to describe statistical significance at the 2-sigma level, corresponding to a confidence level of 95% (which is what we are dealing with), are "very probably." For significance at the 3-sigma level (corresponding to a confidence probability of more than 99.9%) the words "almost certainly" are used. It is not clear where "beyond reasonable doubt" stands in relation to "almost certainly", but is a great deal stronger than "very probably." The phrase is the standard required of the prosecution in most criminal cases within an adversarial system and is the highest level of burden of persuasion.

To suggest that this level of proof has been obtained when $x - 2u > L$ seems to us absurd. We need only imagine a sequence of compliant lots, in which the expected small percentage (e.g. 5%) fails the test. In what sense can the non-compliance of the unfortunate 5% of lots be said to have been "proved beyond reasonable doubt?" We end up with a potentially large collection of lots for each of which we claim 95% certainty of non-compliance, and none of them are in fact non-compliant.

This is a statistical minefield that we see no need to enter. Possibly some countries are required by inappropriate legislation to assert "proof beyond reasonable doubt" before rejecting product on the basis of a compliance test. Codex may have a part to play in rectifying this situation. If a test has rejection rates at various analyte levels that have been agreed to be fair and reasonable, the fact that a lot has failed the test is sufficient reason to reject it. But this presupposes that these rejection probabilities have been considered. In the development of the theory of measurement uncertainty, we would urge that attention be given to this point.

This section seems to imply that an exporting country may be required to certify that $x + U < L$ as part of the export-import procedure. It may not have been appreciated that this would require manufacturers to manufacture product with a true analyte level below $L - 2U$ to have a good chance if its being acceptable for export. This seems unreasonable as it has the potential to impose a significant and unwarranted cost on manufacturers and exporters and might not be controllable.

A manufacturer who almost always has test results $< L$ is entitled to feel confident that almost all his product meets specification, and a test that $x < L$ for each lot may in such cases be quite adequate. Note that if the true level were in fact just below L (where it is quite entitled to be) the probability of a result exceeding L would be nearly 50%.

Situation III suggests a suspicion of non-compliance when there is not strong evidence. Clear conclusions of compliance or non-compliance need to be stated.

Final paragraph, second sentence: It should be made clear that this advice is directed to Commodity Committees. Note that it will be impossible for Commodity Committees to allow for measurement uncertainty without an agreed procedure for its estimation. Also note that, according to Section 4, the measurement uncertainty quoted by different laboratories using the same method of analysis may be different.

Section 9.2

The source of the quote "It is stated …" should be referenced.

Third paragraph: It should be made clear that this advice is directed to Commodity Committees.

Section 10

We agree that this section should be revised to align with the Guidelines on Settling Disputes.