CODEX ALIMENTARIUS COMMISSION E

Food and Agriculture Organization of the United Nations

World Health Organization

Viale delle Terme di Caracalla, 00153 Rome, Italy - Tel: (+39) 06 57051 - Fax: (+39) 06 5705 4593 - E-mail: codex@fao.org - www.codexalimentarius.org

Agenda Item 5 **CX/MAS 14/35/5**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING

Thirty-fifth Session
Budapest, Hungary, 3 - 7 March 2014

DISCUSSION PAPER ON CONSIDERING PROCEDURES FOR ESTABLISHING CRITERIA

(Prepared by the eWG chaired by the United States of America)

BACKGROUND

1. At its 34th session, the Codex Committee on Methods of Analysis and Sampling (CCMAS) Endorsement Working Group considered for endorsement the "Proposed Draft Performance Criteria for Reference and Confirmatory Methods for Marine Biotoxins" (Section I-8.6 Determination of Biotoxins) in the Standard for Live and Raw Bivalve Molluscs (CODEX STAN 292-2008) (para. 99, Appendix VII).

2. CCMAS did not endorse the Criteria, but agreed to the establishment of an Electronic Working Group (eWG) to create a discussion paper considering procedures for establishing criteria for i) multi-analyte methods that are used for specifications that require a combination of components, or use toxic equivalency factors (TEF), and ii) Type I methods. Where there is considerable or statistical overlap between (i) and (ii) these will be considered together (para 49 of REP13/MAS).

3. The Codex Committee of Fish and Fishery Products requested (CL 2013/16-FFP) further comment on the Draft Performance Criteria for Reference and Confirmatory Methods for Marine Biotoxins (Section I-8.6 Determination of Biotoxins) in the *Standard for Live and Raw Bivalve Molluscs* as presented in Appendix VII of REP13/FFP.

4. In response to that request, Norway provided detailed comments and presented an alternative approach for establishing Performance Criteria for Marine Biotoxins (Norwegian Response to CL 2013/16-FFP, 33rd Session CCFFP). The Norwegian Response, which contains a detailed discussion of the applicability of the Horwitz Equation is provided for reference as Annex 2.

[Canada: The information provided in Annex 2, applicability of Horwitz equations outlines the issues for inapplicability in situations where a sum of components is required. Further limitations of the Horwitz equations are elaborated by Dr. Horwitz in J.AOAC Vol 89, No. 4, 2006. On page 1102 of that publication, Dr. Horwitz states that the equations are not applicable to "empirical analytes, i.e., those that are method-dependent, such as moisture, ash, fiber, and similar method-defined analytes, whose composition is ill defined and whose concentration estimate depends on the specific details of the method." This is an important factor when considering options in Part I since the current guidelines for establishing numeric values for rational methods in the Procedural Manual are largely based on the Horwitz functions.]

5. At the 33rd session of CCMAS, a report of the InterAgency Meeting (IAM 24) was presented (para 79 of REP12/MAS, Agenda Item 6, CRD 2). The IAM report contained a discussion paper on the application of the criteria approach to Defining Methods (Type I). Although the IAM report was presented during the 33rd session, there is no record of further comment/consideration of the discussion paper. Many of the concepts presented in that paper are relevant to, and in some cases paraphrased in, this discussion paper. The IAM discussion paper has been provided for reference as Annex 1.

6. Within this discussion paper specific examples (analyte names and concentrations) are presented. The inclusion of these specific examples is to demonstrate and clarify how the described general approach might be used. It is not to present specific criteria for support/endorsement. Therefore, reviews of the discussion should focus on assessing the general approach and not on the acceptability of the specific examples.

7. The *Guidelines for Establishing Numeric Values for the Criteria* listed in the Codex Procedural Manual (21st Ed. 2013, English Version, pp 66) are applicable to methods where the individual concentrations of multiple analytes are determined within a single method (sometimes referred to as multi-analyte methods) and the specification is established for the individual compounds. For example, AOAC Official Method 985.35 can be used to determine a number of elements (Ca, Mn, Fe, etc.) in infant formula. Because the current Criteria Approach is applicable in these cases, they will not be discussed in this document.

[Canada: We agree. The elements determined in such a method are single analyte determinations. Other examples could include multi-residue pesticide methods where each pesticide has its own MRL. Existing Codex guidance and principles would apply.]

8. The eWG had over 50 participants. The list of participants and affiliations are listed at the end of the document, prior to Annex 1.

9. The initial draft of the discussion paper presented a number of options for each of the topics (i.e sum of components, TEF, and Type I). This final version of the discussion paper includes all of the comments made by the participants of the eWG. Each comment is attributed to the specific participant and is captured as italics text in square brackets. In most cases the comments have been added near the specific text on which they comment or at the end of the Section if the comment is more general. A number of participants presented additional options for establishing criteria. These are not captured as comments or attributed to a participant, but are presented as additional options (Options 2-3, 3-3, 3-4). The initial draft avoided making specific recommendations on which approach would be most beneficial and this final document has the same format. However, a number of general recommendations have been added after the Part III conclusions.

INTRODUCTION

The Procedural Manual establishes *General Criteria for the Selection of Methods of Analysis* (21st Ed. 2013, English Version, p 64). Methods are evaluated on the characteristics of selectivity, accuracy, precision, limit of detection, sensitivity, practicability and applicability. It also allows for the establishment of other criteria as required and offers some guidance on choosing between different methods. The Procedural Manual also allows for the "Criteria Approach" as an alternative to the endorsement of a specific method (ibid). The Criteria Approach enables the establishment of a set of criteria (numeric values) which must be met by a method in order for the method to be applicable (i.e. "fit for purpose") to a specific standard. The Criteria Approach is applicable to fully validated Type II and III methods, except for methods such as PCR and ELISA, but it is not applicable to Type I methods. The Criteria Approach currently requires information on Applicability, Minimum Applicable Range, Limit of Detection and Quantitation, Precision (with criteria for reproducibility relative standard deviation), Recovery and Trueness (Procedural Manual 21st Ed.2013, English Version pp 65-77).

Two approaches for establishing criteria have been described in the Procedural Manual. The first utilizes the specified limit (maximum or minimum limit) to establish numeric criteria for the characteristics mentioned above and is summarized in Table 1. The second involves the conversion of a specific method to establish numeric criteria for the parameters listed in Table 1. Although the method should be validated and appropriate for the analyte and commodity, there is not a specific requirement that the method be endorsed prior to being "converted" to criteria. Although it is not specifically stated in the Procedural Manual, the *Guidelines for Establishing Numeric Values for Criteria* were developed considering only single analyte determinations. That is, methods where the concentration of a specific analyte is measured and that determination is assessed against a specification.

Table 1: Guidelines for establishing numeric values for the criteria:

Applicability:	The method has to be applicable for the specified provision, specified commodity and the specified level(s) (maximum and/or minimum) (ML). The minimum applicable range of the method depends on the specified level (ML) to be assessed, and can either be expressed in terms of the reproducibility standard deviation (sR) or in terms of LOD and LOQ.
Minimum applicable range:	For ML ≥ 0.1 mg/kg, [ML - 3 s_R , ML + 3 s_R] For ML < 0.1 mg/kg, [ML - 2 s_R , ML + 2 s_R] s_R[13] = standard deviation of reproducibility
Limit of Detection (LOD):	For ML ≥ 0.1 mg/kg, LOD ≤ ML · 1/10 For ML < 0.1 mg/kg, LOD ≤ ML · 1/5
Limit of Quantification (LOQ):	For ML ≥ 0.1 mg/kg, LOQ ≤ ML · 1/5 For ML < 0.1 mg/kg, LOQ ≤ ML · 2/5

Precision:	For ML ≥ 0.1 mg/kg, HorRat value ≤ 2 For ML < 0.1 mg/kg, the RSD_{TR} < 22%. RSD_R[14] = relative standard deviation of reproducibility. $RSD_R \leq 2$. $PRSD_R$			
Recovery (R):	**Concentration**	**Ratio**	**Unit**	**Recovery (%)**
	100	1	100%(100 g/100g)	98 – 102
	≥10	10^{-1}	≥10% (10 g/100g)	98 – 102
	≥1	10^{-2}	≥1% (1 g/100g)	97 – 103
	≥0.1	10^{-3}	≥0.1% (1 mg/g)	95 – 105
	0.01	10^{-4}	100 mg/kg	90 – 107
	0.001	10^{-5}	10 mg/kg	80 – 110
	0.0001	10^{-6}	1 mg/kg	80 – 110
	0.00001	10^{-7}	100 µg/kg	80 – 110
	0.000001	10^{-8}	10 µg/kg	60 – 115
	0.0000001	10^{-9}	1 µg/kg	40 – 120
Trueness:	Other guidelines are available for expected recovery ranges in specific areas of analysis. In cases where recoveries have been shown to be a function of the matrix other specified requirements may be applied. For the evaluation of trueness preferably certified reference material should be used.			

The criteria in Table 1 must be approved for the determination in question.

In considering the multiple tasks outlined by the eWG's Terms of Reference and the different methods that are to be discussed, it became clear that there are significant connections between the 3 topics (i.e sum of components, TEF, and Type I). However, there are some critical differences that make it cumbersome to discuss them as a single commentary/topic. Therefore, in an attempt to clarify the discussion, the paper is divided into three parts, each focusing on one aspect, but including reference to the other sections when necessary. Finally, the paper initial draft discussion paper was written in an attempt to present an unbiased view on possible approaches for establishing criteria. The goal was to present a starting point for discussion and not to make a recommendation for a particular approach. This final draft attempts to repeat that position, therefore it avoids recommending a specific approach to establishing criteria.

[Netherlands: Indeed I agree that the topics Type I and multiple components are different topics. When it comes to final guidelines or Procedural manual clauses, it would be good to split.]

PART I: Criteria Approach for Type I Methods

Although not explicitly stated in the terms of the electronic working group (para 49 of REP13/MAS), the scope of this paper is to discuss the development of **numeric** criteria to Type I methods, as opposed to

general criteria for Type I methods. General criteria for the selection of methods already exist in the Procedural Manual (pp 64) and are applicable to Type I methods.

[Canada agrees that the general criteria for method selection apply to Type I methods.]

The Procedural Manual defines a Type I Method (Defining Method) as a "*method which determines a value that can only be arrived at in terms of the method per se and serves by definition as the only method for establishing the accepted value of the item measured."* (e.g moisture by loss on drying, salt in brine by density). A Type I method is often referred to as an empirical method, while Type II, III, and IV methods are referred to as "rational methods" and are methods where the measurand can be determined independently of the method (e.g concentration of a Ca in infant formula).

[Canada: Annex I also suggests the term "operationally defined" methods which may be an important distinction between certain type I methods. For example, moisture and ash would be operationally defined entities ie. Whatever is volatile under specific conditions or the amount of residue remaining after ignition under specific conditions. Acid value in oil is also a type I situation in which the amount of acid equivalents is precisely measured but which acids are present is unknown but could be determined by other methods.]

Option 1-1: Apply the current Guidelines for establishing numeric values for Type II/III methods to Type I.

> Many of the characteristics (Table 1) for establishing performance criteria for Type II and III methods could also be generally applicable to Type I methods. For instance, minimum applicable range, precision and even LOD and LOQ could be established as part of the criteria for Type I methods. However, because of the definition of a Type I method, the trueness criterion is not applicable to a Type I method. By definition, if a Type I method is performed correctly it produces the true unbiased result.

[Canada: Again referring to JAOAC 89, 4, 2006, we are not sure that minimum applicable range and precision values can be established using the methodology in the Procedural Manual due to the fact that these are based on Horwitz functions which appear not to be valid in Type I situations. However, noting that the Horwitz functions are based on a generalized model, the use of Horwitz functions may still provide acceptable approximations which can be used. If there could be recognition of the limitations of Horwitz but for expedient and pragmatic considerations be allowed, this option may be viable in some Type I situations.]

[Observer Organization: By definition Type I methods are technique dependent therefore establishment of characteristics like LOQ/LOD is defined by the technique and equipment utilized, while attributes like precision may be related to pure technical conduct and pre-analysis steps like sample preparation and composite sampling. Therefore, Type I methods and newer procedures are better characterized by option two below]

As is described in the IAM 24 Discussion Paper, in many cases the agreement to use a Type I method has the dual effect of defining not just the method, but also the measurand. For example, when a product is submitted for testing by AOAC 925.4 (drying 2 g of product at 95-100 °C to a constant weight) it has been agreed that the weight loss will be considered moisture, but it should be understood that all of the weight loss is not due only to water, but that other volatile compounds have been lost and will contribute to the result. Considering this, it is easy to conceive of developing a method which, when calibrated against a Type I method, produces comparable results. Conversion factors may also be used to calculate a Type I result from a Type II/III method. The toxic equivalency factors (TEFs) developed by Oshima[1] for calculating saxitoxin equivalents are a specific example of such a conversion factor. Based on the concept of "defining the measurand" the goal of "establishing Criteria for Type I methods" could be restated as a goal to establish criteria (procedures) to assess if 2 methods are determining the same measurand.

[Canada: The limits set in the standards may be the real issue. The science available at the time the standards were set may have required the use of type I methods to define the measurand. For example, the PSP toxin limits were established using the type I MBA method since there was essentially no other way to determine the value. The standard is in total saxitoxin equivalents but does not define which toxin analogues contribute to the total. If possible, the limits or measurands in the standards need to be defined by Codex

[1] Oshima, Y. Postcolumn Derivatization Liquid-Chromatographic Method for Paralytic Shellfish Toxins *J. AOAC Int.* 1995, 78, 528-532

committees rather than having methods define the measurand or limit. If the PSP toxin analogues to be included in the total toxin were defined in the standard, it would help in application of the Criteria Approach. Additionally, if the PSP limit could also set limits for each included analogue plus a total, we could apply the existing Codex methodology to establish numeric criteria for the individual components. In STAN 193, the aflatoxin limits are set for total aflatoxin and it has been established that the total is defined as the sum of B1, B2, G1 and G2. There is some information on prevalence and importance of B1 also given. The aflatoxin standard falls short by not actually having limits for each toxin. The method performance criteria in STAN 193 are for total aflatoxin, not for individual toxins, as determined by Horwitz functions (precedent for using these equations in a possibly out of Horwitz scope situation, ie sum of components.). It may be difficult to define type I measurands such as moisture or ash but some other type I analytes may be defined eg. Acid value in oil is the sum of free fatty acids listed in the fatty acid composition tables of the standard (STAN 210 or 83) expressed as equivalents of KOH (these are the units of the standard). This allows the type I titration method to be used to get the equivalents of KOH or a rational GC method that measures each free fatty acid to be used. It still needs a mechanism for multi-component analytes to be developed.

In essence, in type I situations, the method defines the analyte. When considering the establishment of numeric method performance criteria in type I, it could be recommended by CCMAS that, if possible, Commodity Committees define analytes and/or analyte components that are to be considered in the limits set out in standards.]

Option 1-2: Establish guidelines/procedures for comparing the performance of 2 methods.

In this approach it would be necessary to have an endorsed Type I method which could be converted to criteria for precision and minimum applicable range. The endorsed method would then be used to assess the performance of the non-endorsed method. What would need to be determined for this type of comparison is what type and number of samples would be required to establish that the 2 methods' performance is not statistically different. While certified reference materials might assist in this assessment, it would most likely be necessary to analyze a large number of samples across different matrices in order to establish equivalent performance. Realizing that even with an extensive comparison study, there is a possibility that some subset of samples will not have comparable performance on both methods. Once the method has shown equivalency it could be listed in the standard as an alternative method.

[Canada: We agree that showing equivalence between two methods may be a challenge and may require some guidance on how to do this, particularly since the rational method is measuring well defined analytes while the empirical method does not. In essence, the two methods may actually be measuring something different and therefore the results could be statistically different across all or part of the applicable range. For example, crude fat measured by solvent extraction followed by gravimetric determination normally has higher fat values than GC fatty acid methods (AOAC 996.06) particularly at lower fat concentrations due to the increased proportion of other soluble but non-fatty acid components.

It should also be noted that Horwitz in JAOAC 89, 4, 2006 points out that type I methods often have good repeatability (within lab variability) but poor reproducibility (between lab variability). The rational method performance therefore should theoretically always be equal to or better than the empirical method performance. The method performance for the empirical method may not meet the existing Codex requirements but should not have to as the method has been endorsed and is likely in common use. The rational method performance would still need to meet the existing requirements for Codex numerical criteria. Again, a mechanism to deal with the sum of multiple components is needed.]

Because an endorsed Type I method would still be required, this approach would not have the same outcome of removing the endorsement procedure that the Criteria Approach allows with Type II/III methods. However, it could allow for innovation and the development of new methods which have benefits compared to currently endorsed Type I methods.

[Canada: Under this option, there could be both type I and type III methods endorsed. The Procedural Manual would need to be revised accordingly ensure there are no contradictions. This

option seems to be needed to allow for the endorsement of both MBA/RBA and HPLC methods for PSP toxins in the Codex Standard for Bivalve Molluscs.]

Conclusion: There are clearly approaches that could allow for the development of criteria, but it is unclear if they would be beneficial to Codex members or create confusion. Therefore, perhaps the appropriate question should not be **if** criteria can be established for Type I methods, but **should** criteria be established for Type I methods. In considering these questions, recall that Codex methods are not required, except in the case of disputes. Therefore, if a "better" (i.e. faster, cheaper) method with comparable performance to a Type I method is available; an analyst has the option of using that method for routine analysis.

[Australia: It would be our contention that Type I methods are in-place as a "fix" for a historical legacy or atypical analytical issues. Thus while their presence was/is necessary, the objective would be to "grandfather" these methods and replace as appropriate "rational" methods became available. Unfortunately, by defining Type I methods as unique, i.e. the only method from which a measurand result can be derived; we have effectively reinforced their status as beyond replacement. Thus the instigation of Type 1 method equivalency guideline/procedure, while not fully resolving this issue, may provide an avenue to support introduction of alternative methods, if a rational method cannot be found. Further, this discussion should highlight the need for Codex to replace endorsed empirical with rational method criteria and hopefully discourage the generation of Type I 'results' from Type II/III methods.

In addition while certified reference materials are mentioned in the text, we believe they should have a far more significant role in demonstrating method equivalency. Thus page 3 Option 1-2: 4th sentence should be amended as follows:

~~*While certified reference materials might assist in this assessment,*~~ *It would most likely be necessary to analyse a large number of samples across different matrices in order to establish equivalent performance and where available a certified reference material must be included in this assessment.]*

[Canada: CCFFP is looking for a way to establish numerical criteria that can be applied to empirical and rational methods together in order to allow for acceptability of mouse or receptor binding bioassays (defining) and instrumental (rational) methods in official control programs that are acceptable between trading partners. We agree that for routine analysis, use of Codex methods are not really a requirement. However, CCMAS has been presented with proposed criteria from CCFFP. Given the complexity of trying to establish a procedure to elaborate criteria for type I situations, perhaps another option to consider is to treat each on a case by case basis. We understand that this could lead to a patchwork of approaches rather than a single, systematic, orderly approach. A single approach may be very complex and subject to frequent revisions as unforeseen issues arise with particular type I methods.]

[Netherlands: The way you describe the problem of Type I methods gives me the feeling that criteria may be applicable to Type II methods, shown statistically equivalent to the Type I. I can imagine that the bias can be way smaller than the measurement uncertainty. Is it useful to add such a remark? I fully agree that a conversion factor shouldn't convert a physical type II/III entity in a Type I. already for a long time I've seen that as flaw in the system. Maybe this is the time to correct.]

[Norway: Norway fully agrees with the conclusion made by US. Maybe more focus should be on when methods have comparable performance. Some methods classified as a type I methods might be classified as type II/III methods if the conversion factor/calculation was included in a standard and not in the method itself. It might be time to revise the description of type I methods.]

*[United Kingdom: The UK agrees with the conclusion given on page 4 that, "perhaps the appropriate question should not be **if** criteria can be established for Type I methods, but **should** criteria be established for Type I methods". The general question being asked needs to be clarified before any further work is undertaken to develop method performance criteria for Type I methods.]*

[AACC International: We believe there are several options to set criteria for Type I methods. The most important thing from the perspective of AACCI, as an organization that develops and manages standards and approved methods, is to make the guidelines straightforward to understand across analytes and matrices and to use internationally.

We'd like to suggest that a helpful approach would be to sort the methods into like-kind. For example the Type 1 methods listed in Std 234 are mostly proximate assays. The following categories: Protein, Moisture/solids, Fats, Ash and Fiber cover most of these methods. We also recognize that there are some

methods that might prove difficult "to group" for example "drained weight of canned tomatoes" that may need special attention.

There are already publications covering some of these "groups", J. DeVries and P. Wehling proposed criteria for Fiber methods – and a recent manuscript (submitted to JAOAC) by the same authors covers similar work for Moisture/solids methods. They have plans for a further discussion on fat methods.

Experience shows that it is highly unlikely that Horwitz-based criteria can be applied to Type I analyses, but a reasonable approach would be to target key method types and work on establishing criteria based on historical validation data, as per the statistical techniques of Wehling and DeVries.

In addition, there may be some key areas to investigate, such as ELISA methods for gluten or allergens which may warrant deeper study but we believe protein, fat and ash are definitely workable with a similar approach.]

*[Association Of European Coeliac Societies: We support the Conclusion on page 4 "... the appropriate question should not be **if** criteria can be established for Type I methods, but **should** criteria be established for Type I methods". This question should be answered before any further consideration on this item.]*

[Observer Organization: More importantly, there should be a path for innovative, efficient methodology to replace current reference methods that enables data of similar quality to be accepted globally.]

[Dr. Roger Wood: I am wholeheartedly in favour of the conclusion:

*Conclusion: There are clearly approaches that could allow for the development of criteria, but it is unclear if they would be beneficial to Codex members or create confusion. Therefore, perhaps the appropriate question should not be **if** criteria can be established for Type I methods, but **should** criteria be established for Type I methods. In considering these questions, recall that Codex methods are not required, except in the case of disputes. Therefore, if a "better" (i.e. faster, cheaper) method with comparable performance to a Type I method is available; an analyst has the option of using that method for routine analysis.*

To go any further will lead to confusion, I think, and will open the door to deliberate mis-use. But it is clear that there are now some standards produced by ISO/IDF (and it is mainly in the milk sector from memory) that talk about calibrating a routine method against a reference method. And again from memory I am pretty sure that they are for empirical determinations.

An alternative may be to consider whether general instructions could be prepared for comparison calibration of empirical methods for the methods which they produce.]

[Dr. Steve Ellison: On the utility of criteria for type I, though, and on the possibility for numerical criteria, I have a couple of thoughts to throw into the mix.

First, it will clearly not be possible to do away entirely with a decision process for Type I measurands because these methods are intended to give a relative indication of some real-world behaviour and there is an essential choice to be made about what measure to use for each particular purpose.

This consideration alone might make the entire debate moot; if we need a committee decision anyway, what is the benefit of criteria?

However, when choosing between such measures, perhaps with knowledge of the performance of methods intending to implement them, numerical criteria will assist in comparing the different choices and in weeding out very poor possible choices. That is likely to be helpful even if it does not entirely remove the necessity for some choice.

Turning to the kind of criterion, that is quite hard as not all Type I measurands are mass fractions at all. Where they are not, CCMAS or the commodity committee would have no general basis to go on, though that makes them no worse off than now in such cases.

Where the Type I measurand is reasonably expressed as a mass fraction, however, Horwitz might still be applicable in setting boundary conditions.]

PART II: Specification Requires a Combination of Components

For certain commodities or analytes there are specifications where the individual concentrations of multiple analytes are determined by a single method, but the specification is a limit based on the sum of the individual components to determine a Total Concentration. One example of this approach is the determination of

aflatoxins in nuts in Codex Standard 193-2005. The specification is for the concentration of Total Aflatoxin, which is determined as the sum of B1, B2, G1, and G2. As is stated above, the current Criteria Approach in the Procedural Manual was not developed considering specifications which use a "sum of components." There are a number of possible options, each with benefits and drawbacks for establishing criteria in these situations.

Option 2-1: Use the specification (sum of components) as the specified level (maximum/minimum limit) and develop numeric criteria based on this limit and the parameters listed in Table 1.

This approach would follow and be consistent with the current Procedural Manual Guidelines for establishing criteria. Additionally, the numeric criteria that are established through this procedure would be directly related to the specification. This approach of setting criteria based on the specification has been used for aflatoxins in nuts in *The Codex General Standard for Contaminants and Toxins in Food and Feed (*Codex Standard 193-2005).

Notwithstanding the precedent of using this approach in Codex Standard 193-1995, there are a number of questions/concerns which arise from using the "sum of components" for establishing numeric criteria. First, because the Horwitz/Thompson Equation was originally derived based on data associated with individual analytes, it is not directly applicable to determining the predicted relative standard deviation of a "sum of components." Therefore, the Horwitz/Thompson Equation or HorRat cannot be used to establish a numeric value for the Precision. If one were to attempt to apply the Horwitz/Thompson Equation to the "sum of components" it could produce a situation where the precision of one or more individual component would need to exceed 100%. A specific example related to the issues of attempting to use the Horwitz Equation was presented by Norway at the Codex Committee on Fish and Fishery Products (CCFFP) and is included in Annex 2. Although this example uses toxic equivalency factors, the statistical discussion also applies to "sum of component" situations.

[Canada: We agree that using Horwitz functions is not valid for a sum of components. This approach is used in STAN 193 for aflatoxin total (sum of B1, B2, G1 and G2) and we are not certain of the rationale that was used to substantiate it. Perhaps if the drafters of these criteria have a rationale, it could be shared and considered by this WG. In reviewing aflatoxin method validation data for several methods, we note that it is customary for the method performance (repeatability, reproducibility, etc.) to be given for the total sum of components rather than by components. This could be the rationale for the aflatoxin criteria decisions. We do not recommend reopening or debating the aflatoxin criteria. We may wish to consider this a precedent that could allow for the use of Horwitz equations in similar situations. We also recommend documenting the reasons for allowing this "out of scope" use of Horwitz functions.]

A second concern related to developing criteria based on the "sum of the components," is that there are no criteria/requirements related to the detection of any of the individual analytes. Therefore, it would be possible for a method to satisfy the numeric criteria even without the ability to measure one or more of the individual components. In the aflatoxin example the maximum limit for peanuts is 15 □g/kg (Codex Standard 193-1995). If numeric criteria were based solely on this limit, it would be possible to develop a method which meets the criteria but does not detect G2, for example. [Clearly, in the case of aflatoxins, where there are only 4 components this is an unlikely scenario. However, as the number of individual components grows or if the presence of certain components differs based on geographic origin or temporal cycles, it is quite possible that a method could have unacceptable performance for one or more of the individual components.

[Canada: We are not sure that a method that does not detect G2 would be considered as fit for purpose since the residue definition is the sum of all 4 toxins. As such, any fit for purpose method must include G2 in its scope of analytes. To elaborate numeric criteria, we feel that identification of all the components needs to be included in the standard definition of the limit when it includes a sum of multiple components.]

[Observer Organization: (Commenting on Option 1) This approached assumes that all analytes are created equally, ignoring the biological relevance of the individual components based upon historical specifications.]

Option 2-2: Choose a suitable method and convert it into criteria using the guidelines currently listed in the Procedural Manual.

Option 2-2A: The numeric criteria are established from the approved method for each of the individual components.

The approach of converting methods has been approved by CCMAS and is clearly stated as an option for establishing criteria in the Procedural Manual. Because this approach would set criteria for each of the individual components, it would address both of the concerns in Option 1. First, the Horwitz Equation would be directly applicable to the individual components, so a precision value for each component could be established. *[Canada: Is a limit for each component not needed to do this?]* Second, by establishing criteria for each of the components it would address the concern of allowing methods which do not detect all of the individual components. Additionally, by using an established method, the converted values would be realistic levels for the determination of each of the components. One difficulty with this approach is that if the precision requirements for each individual component are established using the guidelines in the procedural manual, the precision on the summation grows smaller as the number of components grows larger (Annex 2).

[Canada: For aflatoxin, it is fairly straightforward as a sum of 4 known or defined components each with a TEF of 1. To assign criteria for each component using the current Codex procedure, we need a limit for each component or we need a way to prorate/divide the total aflatoxin limit over the 4 components to have a value to use as a limit for calculating the individual numeric values. For example, how would one prorate the 15 ug/kg total limit over the 4 components to apply the Horwitz calculations for each component? Divide the ML for total by the number of components? For aflatoxin, this would mean that the precision for each component would be calculated by Horwitz using a limit of 15/4 ug/kg or 3.75 ug/kg.]

A general question about this approach is whether it is "permitted" within Codex to establish criteria for analytes that do not have associated specifications? With this approach, the criteria are only indirectly linked to the specification, however, if the criteria were applicable to the individual components they should be applicable to the sum of the components.

[Finland: PART II: (Specification requires a combination of components)/Option 2-2A: I agree that the numeric criteria are established from the approved method for each of the individual components. In this manner it's able to assure the performance of the analysis for all the components. However, after you have calculated all the parameters (table 1) for individual components and you should sum up/combine the results, it's not that easy. For example, determining LOD and LOQ for the sum of the components, is there a clear rule for that? In addition the other parameters might be difficult to determine for the sum?]

Option 2-2B: The numeric criteria are established based on the specification and on the method performance for individual components.

This approach would utilize the specification and Table 1 to establish certain numeric values for the criteria, but also use a suitable method to establish the criteria for the detection of each individual component. For example, the Minimum Applicable Range, Recovery and Trueness could all be established based on the specification (sum of the components), but the precision would be determined for each of the individual components based on the conversion of an applicable method.

This approach is very similar to Option 2-2A, and would still establish criteria that are not associated with the specification, and it is unclear if this is "permitted." As with Option 2-2A, this approach avoids the application of the Horwitz Equation to a sum, but unlike Option 2-2A, it still establishes numeric criteria directly related to the specification.

Option 2-3: Numeric criteria established based on the ML and the number of components.

This option refers to multi-analyte methods where ML's are not established for each of the components.

The Codex guidelines for establishing numeric values for LOQ are as follows:

Limit of Quantification (LOQ):	For ML ≥ 0.1 mg/kg, LOQ ≤ ML · 1/5 For ML < 0.1 mg/kg, LOQ ≤ ML · 2/5

This is valid for analysing one component. When the ML is based on a sum of components, the LOQ for the individual component should be correspondingly low. When summing two components, the LOQ for each component should be the half for each component, and if summing three components; the LOQ for each component should be 1/3 of the LOQ.

Based on this, the following criteria for LOQ are suggested:

Limit of Quantification (LOQ):	For ML ≥ 0.1 mg/kg, LOQ ≤ ML · 1/5· 1/n For ML < 0.1 mg/kg, LOQ ≤ ML · 2/5· 1/n Where n = number of components

For multi-analyte analyses where all components are weighted equal, n is the number of components/analytes. The criteria for multi-analyte (and single analyte, n=1) would then be as given in Table 1.

Table 1: Numeric values for method performance criteria

Applicability:	The method has to be applicable for the specified provision, specified commodity and the specified level(s) (maximum and/or minimum) (ML).			
Minimum applicable range for the individual components :	For ML/n ≥ 0.1 mg/kg, [ML/n - 3 s_R , ML/a + 3 s_R]* For ML/n < 0.1 mg/kg, [ML/n - 2 s_R , ML/a + 2 s_R] s_R = standard deviation of reproducibility Note that the upper level is above ML for the individual components.			
Limit of Detection (LOD) for the individual components:	For ML/n ≥ 0.1 mg/kg, LOD ≤ ML/n · 1/10 For ML/n < 0.1 mg/kg, LOD ≤ ML/n · 1/5			
Limit of Quantification (LOQ) for the individual components:	For ML/n ≥ 0.1 mg/kg, LOQ ≤ ML /n· 1/5 For ML/n < 0.1 mg/kg, LOQ ≤ ML/n · 2/5			
Precision for the individual components:	For ML/n ≥ 0.1 mg/kg, HorRat value ≤ 2 For ML/n < 0.1 mg/kg, the RSD_R < 44%. RSD_R[14] = relative standard deviation of reproducibility. $RSD_R \leq 2.\ PRSD_R$			
Recovery (R):	**Concentration**	**Ratio**	**Unit**	**Recovery (%)**
	100	1	100% (100 g/100g)	98 – 102
	≥10	10^{-1}	≥10% (10 g/100g)	98 – 102
	≥1	10^{-2}	≥1% (1 g/100g)	97 – 103
	≥0.1	10^{-3}	≥0.1% (1 mg/g)	95 – 105
	0.01	10^{-4}	100 mg/kg	90 – 107
	0.001	10^{-5}	10 mg/kg	80 – 110
	0.0001	10^{-6}	1 mg/kg	80 – 110
	0.00001	10^{-7}	100 µg/kg	80 – 110
	0.000001	10^{-8}	10 µg/kg	60 – 115
	0.0000001	10^{-9}	1 µg/kg	40 – 120
Trueness:	Other guidelines are available for expected recovery ranges in specific areas of analysis. In cases where recoveries have been shown to be a function of the matrix other specified requirements may be applied. For the evaluation of trueness preferably certified reference material should be used.			

** For multi-analyte analyses where all components are weighted equal, n= a= number of components/analytes*

Example of option 2-3, multi-analyte :

Aflatoxin, consisting of 4 analytes, B1, B2, G1 and G2, in peanuts.

The ML = 15 µg/kg ,

As there are 4 analytes, n = 4,

$ML/n = \frac{0.015}{4} mg/kg = 0.00375\ mg/kg$

Using the excel spreadsheet on www.nmkl.org under "how to get method criteria based on ML", the following are established:

Minimum applicable range for the individual components :	0.002*– 0.022** mg/kg = 2 – 22 µg/kg * corresponding to ML = 0.00375 mg/kg **corresponding to ML = 0.015 mg/kg
Limit of Detection (LOD) for the individual components:	0.00075 mg/kg = 0.75 µg/kg
Limit of Quantification(LOQ) for the individual components:	0.0015 mg/kg = 1.5 µg/kg
Precision for the individual components:	$RSD_R < 44\%$
Recovery (R):	40-120%

Examples on methods fulfilling the criteria:

AOAC 999.07 Immunoaffinity Column LX with post column derivatization

AOAC 2005.08 LC with Post-column photochemical derivatization

Examples on methods not fulfilling the criteria:

AOAC 975.36 (Romer minicolumn method) applicable for ≥ 10 µg/kg

AOAC 990.34 (Enzyme Linked Immunosorbent (ImmunoDot Screen Cup) Screening Assay ≥ 20 µg/kg

AOCS –AOAC 970.45, AOCS –AOAC 998.03. AOAC 993.17 Thin Layer Chromatography

Conclusion: There are a number of options for establishing criteria for specifications where a "sum of components" is used. Given that there are already specifications which use the "sum of components," there is a clear need for establishing guidelines. One difficulty may be in establishing guidelines which are generally applicable to all of the situations where sums of components are used.

[Observer Organization: Why not separate analysis from interpretations? Method performance standards enable the generation of quality information that can then be summarized via summation and comparison to biologically relevant specifications.]

PART III: Specification Requiring the Use of Toxic Equivalency Factors (TEFs)

For certain commodities or analytes there are specifications where the individual concentrations of multiple analytes are determined by a single method, the concentrations are converted to a "toxic equivalence" using a toxic equivalency factor (TEF) and the specification is a limit based on the sum of equivalencies. One example of this approach is the determination of the saxitoxin group in the *Standard for Live and Raw Bivalve Molluscs* (Codex Standard 292-2008). The specification is for the concentration of saxitoxin equivalents which is determined from 12[2] saxitoxin congeners each multiplied by a TEF and summed. TEFs are also used in other determinations, such as dioxins and dioxin-like PCBs, and PAHs. As is stated above, the current Criteria Approach in the Procedural Manual was not developed considering specifications which use TEF or a sum of toxic equivalents. The options and discussion presented below are very similar to those presented for the "sum of components," however the use of TEFs does present some particular challenges.

The use of a TEF to determine a "toxic equivalent" requires a calculation, and if this calculation is part of the method, then historically CCMAS would consider such methods as Type I. Even if the analytical procedure to determine the value prior to conversion was rational (Type II/III), the final determination is Type I because the calculation is empirical. A possible alternative to including the TEFs in the method would be to

[2] There are more than 12 saxitoxin congeners identifed, however the currently endorsed method (AOAC 2005.05) only lists 12 compounds.

include them in the standard. This could be done directly with a table which lists the analytes and corresponding TEFs or through reference to TEFs which are evaluated and updated regularly.[3]

[Observer Organization: Yes – completely separate analysis methods from TEF's, methods describe how to generate quality data (accurate/precise, specific....). Data interpretation or reduction is a separate topic.]

Option 3-1: Use the specification (sum of equivalencies) as the specified level (maximum limit) and develop numeric criteria based on this level and the parameters listed in Table 1.

As with the "sum of components" discussion above there are some concerns about applying the Criteria Approach in the Procedural Manual to methods which use TEFs and sum of components. The Horwitz/Thompson Equation are not applicable in these cases. The Horwitz Equation for methods which use both a sum of components and TEFs has not been evaluated or established. The Horwitz Equation was originally derived based on data associated with individual analytes and TEFs were not used. Additionally, depending on the number of individual components, if the Horwitz Equation is used to establish the precision for the sum, then in certain cases the precision on each of the components can rise to >100%. (A specific example related to the issues of attempting to use the Horwitz Equation was presented by Norway at CCFFP and is included in Annex 1). Second, by developing criteria based on the "sum of the components," there are no criteria related to the detection of any of the individual analytes. Therefore, it would be possible for a method to meet the criteria (i.e. be applicable) without the ability to measure one or more of the individual components.

[Canada: The issue for PSP is similar to the aflatoxin situation but complicated by the fact that there needs to be a TEF consideration for each component in order to report in saxitoxin toxicological equivalents. Does one divide the total by each component adjusted for TEF? Which TEFs are applied? The TEFs must be identified and agreed upon by the responsible Codex committee. The individual components that are to be included in the sum and the TEFs need to be agreed upon and listed in the Standard (or other reference such as the FAO/Codex/WHO website)]

Option 3-2: Choose a suitable method and convert it into criteria using the guidelines currently listed in the Procedural Manual.

Option 3-2A: The numeric criteria are established from the approved method for each of the individual components.

The approach of converting methods has been approved by CCMAS and is clearly stated as an option for establishing criteria in the Procedural Manual. Because this approach would set criteria for each of the individual components, it would address both of the concerns in Option 1. First, the Horwitz Equation would be directly applicable to the individual components, so a precision value for each component could be established. Second, by establishing criteria for each of the components, it would address the concern of allowing methods which do not detect all of the individual components. Additionally, by using an established method, the converted values would be realistic levels for the determination of each of the components. One difficulty with this approach is that if the precision requirements for each individual component are established using the guidelines in the Procedural Manual, the precision in the summation grows smaller as the number of components grows larger (Annex 2).

[Canada: The issues are the same as for Part II, Option 2-2]

[Observer Organization: The quality of data from each individual component measured in a multi-component analysis should be addressed independently to describe method performance. This enables each component to be validated to be appropriately extracted, detected, and quantitated in the target matrix ensuring quality data is available for assessment. Trying to relate the performance of one component to another in terms of precision is meaningful as presented here]

A general question about this approach is whether it is "permitted" within Codex to establish criteria for analytes that do not have associated specifications? With this approach the criteria are only indirectly linked to the specification, however if the criteria were applicable to the individual components it should be applicable to the sum of the components.

[3] The World Health Organization (WHO) maintains and assesses the TEFs for dioxin and dioxin-like PCBs on a regular schedule.

[Observer Organization: Why not completely divorce analysis from conclusion? 1. Define what the required performance requirements for the methodology for the matrix and analyte combination. 2. Define specifications (sums using TEF or other factors) based upon biological and toxicological relevance]

Option 3-2B: The numeric criteria are established based on the specification and on the method performance for individual components.

This approach would utilize the specification and Table 1 to establish certain numeric values for the criteria, but also use a suitable method to establish the criteria for the detection of each individual component. For example, the Minimum Applicable Range or Limit of Detection and Limit of Quantitation, Recovery and Trueness could all be established based on the specification (sum of the components), but the precision would be determined for each of the individual components based on conversion of an applicable method.

This approach is very similar to Option 3-2A and would still establish criteria not associated with the specification, and it is unclear if this is "permitted." As with Option 3-2A, this approach avoids the application of the Horwitz Equation to a sum, but unlike Option 3-2A, it still establishes numeric criteria directly related to the specification.

For both Option 3-2A and 3-2B, the procedure for conversion would need to address if and when TEFs would be used in establishing numeric criteria. For instance, if the LOQ is utilized for establishing the minimum applicable range, should the LOQ calculation factor in the TEF (i.e divide the concentration by the TEF) for each analyte or base the calculation of LOQ solely on the concentration determined during the method validation?

[Canada: The proposal drafted by Norway for marine toxin criteria has been elaborated along these lines and may be a useful example to have available for discussion. They have put a lot of effort into the work which should be recognized. Using the minimum lowest validated levels with suitable performance for each component appears to be a sound methodology for conversion of methods.]

Option 3-3: Establish the applicable range using the limit set out in the Standard (recognizing that Horwitz does not apply to sums of components but allowing use for pragmatic reasons).

For example, for PSP, ML is 0.8 mg/kg total STX equivalents. Using the Procedural Manual method with Horwitz, the minimum is calculated as 0.4 mg/kg total. The components and TEFs are agreed upon by the Commodity Committee. Candidate methods must be validated as per the Codex general method selection criteria. The candidate method includes all the required components and has been suitably validated at the calculated minimum total level of 0.4 mg/kg. Type I candidate methods are assumed to include all the components and need to meet the minimum requirement. For rational multi-component methods, the sum of all component LOQs (TEF adjusted) must be ≤ the minimum total of 0.4 mg/kg. It is understood that there are different ways to determine LOQ which can lead to lower values. This is a major constraint of this proposed approach. It does however, allow for both type I and rational methods to be considered and it does use the limit as established in the Standard.

Option 3-4: Consider the toxicity of the individual components when establishing criteria

If the components involved are not considered to be equally toxic, and toxicological equivalent factors (TEFs) are to be taken into account for checking compliance with the maximum limit (ML), this should be reflected in the requirements. For instance if a component A is 10 times more toxic than component B, the LOQ for component A should be about 10 times lower than the LOQ of component B. The factor n, could then be expressed as:

$$n = TEF(x_i) \cdot \sum_{i=1}^{n} TEF(x_i)$$

Where

$TEF(x_i)$ is the toxicological factor for the indivisual component x and

$\sum_{i=1}^{n} TEF(x_i)$ is the sum of the toxicological factors for all the individual components.

If the all components are equally toxic, TEF = 1, then the sum of TEFs, would be equal to the number of components (multi-element without TEFs). The criteria are estimated according to Table 1 above, except for the upper level of minimum applicable level. For the upper level, the ML is divided with the TEF value for the individual component (TEF(x))

Minimum applicable range for <u>the individual components</u> :	For ML/n ≥ 0.1 mg/kg, [ML/n - 3 s_R , ML/a + 3 s_R] For ML/n < 0.1 mg/kg, [ML/n - 2 s_R , ML/a + 2 s_R] s_R = standard deviation of reproducibility Note that the upper level is above ML for the individual components.

$$n = TEF(x_i) \cdot \sum_{i=1}^{n} TEF(x_i) \quad and \quad a = TEF(x)$$

Example of option 3, multi-analyte including toxicity equivalent factors (TEFs)

Total toxicity of Okadaic acid

The ML = 0.16 mg/kg OA equivalents,

As there are 3 analytes, with TEFs (a) of 0.5 and 1, respectively, the values for n are given in the table below.

PSP TOXIN	**TEFs *a***	***n***	**ML/n (mg/kg)**	**Min appl level (mg/kg)**	**LOD (mg/kg)**	**LOQ (mg/kg)**	**RSD_R (%)**	**R (%)**
OA	1	1 · 2.5=2.5	0.064	0.036* - 0.26**	0.01	0.03	44	60-115
DTX1	1	2.5	0.064	0.036* - 0.26**	0.01	0.03	44	60-115
DTX2	0.5	0.5 · 2.5=1.25	0.128	0.044* - 0.32**	0.01	0.03	44	60-115
	Sum 2.5							

* ML in the Excel spreadsheet is = $\dfrac{ML}{TEF(x_i) \cdot \sum_{i=1}^{n} TEF(x_i)}$

** ML in the Excel spreadsheet is = $\dfrac{ML}{TEF(x_i)}$

Methods meeting the criteria:

- "EU-Harmonised Standard Operating Procedure for determination of Lipophilic marine biotoxins in molluscs by LC-MS/MS"
- EU harmonized standard operation procedure for determination of OA Group toxins by LC-MS/MS[2]
- EN 16204:2012 LC-MS/MS

There might also be typical type I methods (such as bioassay methods) for the determination of biotoxins which are preferred used by many countries due to cost savings. If it can be proved that a defining method provides equivalent results as rational methods that are fulfilling the criteria, the defining method should be considered fit for the purpose. For proving equivalency, results from appropriate proficiency testing schemes and/or method comparison studies could be reviewed. The studies have to be conducted according to international protocols and reviewed by independent third party. If the comparing results, including the uncertainty of the results, are overlapping the methods could be considered to provide equivalent results. This approach could also be applicable in the choice of an alternative method to a specified Type I method.

Conclusion: There are a number of options for establishing criteria for standards where a toxic equivalent is used as the specification. Based on the request from CCFFP, there is a need to develop guidelines. One difficulty may be establishing guidelines which are generally applicable to all of the situations where TEFs are used.

[Australia: As you can see from our comments below, we believe these two "parts" could be merged and

dealt with in a similar manner.

We would prefer the current CCMAS criteria approach is retained or wherever necessary only modified to enhance the coverage to multi-analyte methods. To support this, the application of the Horiwtz/Thompson Equation should be restricted to individual analyte measurements, not to 'summations' or values where a factor has been applied e.g. TEFs. It would be our recommendation that the method of summation and application of a TEF to the analytical result should be specified in the relevant Codex Standard but not in the analytical method.

In a multi-component analysis, the individual analyte measurements are correlated and not independent as we believe is implied in Annex 2. If the measurements were 'independent' we could expect a high 'cancelling-out' effect from the individual positive/negative deviations from the "actual" and reduced overall MU for the 'total analyte concentration summation' relative to the 'individual analyte concentrations'. Generally, the %RSD or MU of an individual analyte concentration from a multi-analyte analysis should not be directly compared to the %RSD or MU of a 'total analyte concentration summation' to assess performance.]

[Japan: With regards to setting criteria for multi-analyte method, combination of components and use of TEFs, we think it better to excluding the TEFs from analytical method because the establishment of TEFs is not influenced by analytical procedure. TEFs are established by Risk Assessment body evaluating toxicological data and can be re-evaluated and updated using latest scientific data.

The value of TEFs is not directory related to performance of analytical methods. If one of analyte having high TEFs value is needed to be analyzed in lower concentration, the concentration should be clearly shown to set the suitable criteria. The target concentration should be calculated by Codex Committees using the toxicological data, occurrence data and maximum level of the target analyte.

If analytical methods include the formula to calculate the value of total toxicity, it is better only to refer the use of the latest TEFs rather than stipulating values of TEFs in the method.]

[Netherlands: With respect the multiple component/TEF case I would like to support case 2a and would be in favour of expressing that in the paper. The criteria to the individual analytes are straight forward. The criteria to the sum (case 2b) are misty as usually the uncertainties in the individual components are not independent. In my view Reproducibility standard deviations cannot be summed up in the way the Norwegian paper does, because of the correlation. Taking this correlation argument it may be useful to reiterate on option 1. The extreme correlation case are multicomponent pesticides with fixed abundances during production. In these cases it may be possible to apply standard Horrat-analysis, with additional requirements like minimum LODs to be sure that all relevant components as detected, like the afla g2.]

[Norway: Norway thinks that the suggestion made by NMKL on the 31st of January is a possible way forward. This will depend on if the TEF-factors are to be included in the method or not. One challenge is that for mycotoxins CCFFP wants the bioassay to be a Codex method and this method will be a type I method.]

[United Kingdom: With respect to Part II of the discussion paper, the UK welcomes the work undertaken by the USA and Norway in relation to method performance criteria but suggests that issues relating to toxic equivalent factors and analytical method performance criteria should be kept separate. Within some analytical areas (e.g. PSP shellfish toxins) discussions on suitable toxic equivalent factors to be employed are still ongoing hence the UK considers it prudent to keep analytical issues and post-analysis data manipulation (i.e. application of toxic equivalent factors) separate. The UK recognises the comments made within the Discussion document that strictly speaking the Horwitz function should not be used to estimate precision for specifications that require a combination of components but a pragmatic approach would be to set criteria in such specifications on an individual analyte basis rather than be related to toxic equivalent factors or the combined value set down. Where TEFs are set these normally relate to the most toxic substance within the suite of substances being analysed so a pragmatic approach would be to ensure that

any analyte specific method performance criteria stipulated are based upon the most toxic substance itself (e.g. for PSP shellfish toxins, STX, Oshima) and therefore be applicable to all other analytes/components determined within the specification.]

[Editorial Note The underlined portion in the comments below by Dr. Ellison have been presented previously, at the end of Part I. However, because the latter statements addressing TEFs depend on the former I have included them again here to give proper context.]

[Dr. Steve Ellison: On the utility of criteria for type I, though, and on the possibility for numerical criteria, I have a couiple of thoughts to throw into the mix.

First, it will clearly not be possible to do away entirely with a decision process for Type I measurands because these methods are intended to give a relative indication of some real-world behaviour and there is an essential choice to be made about what measure to use for each particular purpose.

This consideration alone might make the entire debate moot; if we need a committee decision anyway, what is the benefit of criteria?

However, when choosing between such measures, perhaps with knowledge of the performance of methods intending to implement them, numerical criteria will assist in comparing the different choices and in weeding out very poor possible choices. That is likely to be helpful even if it does not entirely remove the necessity for some choice.

Turning to the kind of criterion, that is quite hard as not all Type I measurands are mass fractions at all. Where they are not, CCMAS or the commodity committee would have no general basis to go on, though that makes them no worse off than now in such cases.

Where the Type I measurand is reasonably expressed as a mass fraction, however, Horwitz might still be applicable in setting boundary conditions.

For example, in the case of toxic equivalent, we are essentially setting a limit in terms of the equivalent mass fraction of some (usually) particularly toxic species. Since it is a mass fraction, Horwitz should be informative.

It happens that for toxic equivalent we are also almost always dealing with substances that are off the Horwitz range, taking you into the Thompson limit of about 22%RSD for useful methods. So if you are talking about toxic equivalents below the 0.1ppm level, any method that does as well as 22% for the toxic equivalent ought to be acceptable. For higher levels, the Horwitz value is likely to be the criterion.

Whatever precision criterion is chosen in a 'toxic equivalent' case, though, I strongly suspect that the key thing is to ensure that it is applied to performance for the most toxic individual component, as this will be at the lowest mass fraction and therefore have the poorest expected precision. If Horwitz's model applies ideally, all the others would be more precise if they were the dominant component in a test sample at the limit because their mass fractions would be higher. It follows that if a method works when the most toxic component is at the toxic equivalent limit, it should work better when any other weighted sum is at the limit. This kind of argument would allow a comparatively simple criterion that need only be demonstrated in a worst case (though I would recommend additional validation for 'typical' samples, as mass fraction is very far from the only thing that affects precision!).

Similar arguments apply to a simple sum of a finite number of components (eg aflatoxins B1, B2, G1 and G2.); the best predicted precision occurs when only one is present and at the limit, in which case Horwitz applies. If the method achieves that precision on total mass fraction for the most difficult case (which my intuition suggests is the case when all components are present at equal mass fraction, though that would need demonstrating) then you would have a method that would not perform worse than the horwitz prediction.

So the idea common to these (mass sums and (linear) toxic equivalent calculations) would be to apply the Horwitz-Thompson criterion to the total mass fraction _in the worst case_, and perhaps do some research to make sure we are clear on what that worst case is.

On Type III (if I have the 'permitted equivalents' correct) the criteria are fairly straightforward - any method that, within its defined scope, is unbiased compared to the Type I or Type II method and has equal or better precision should be acceptable. Just make sure it's checked on a representative range of sample types; one

strength of Type I 'methods' is that they effectively define sample type out of the problem, while ostensibly equivalent methods may well show strong effects due to sample type.]

RECOMMENDATIONS

Based on the number of participants in the eWG there is clearly great interest in the concept of developing criteria for Type I methods and/or for multi-analyte methods where a sum of components or TEFs are utilized. The draft discussion paper was presented as starting point for discussion and therefore, did not attempt to reach consensus on a particular approach. However, based on the comments there are a number of general recommendations that should be consider by CCMAS.

1. The establishment of Criteria for the different circumstances (Type I and multi-analyte method) should be addressed separately both during development of the criteria and within the Procedural Manual.
2. With respect to Type I methods, CCMAS should address the following questions before there are additional discussion on the possible approaches:
   - If criteria for Type I methods should be established, or
   - if a procedure for determining when methods have comparable performance should be developed, or
   - if the current system should remain unchanged.
3. CCMAS should work to establish a Criteria Approach for multi-analyte methods. If possible a single approach for "sum of components" and TEFs should be established, but it may be necessary to address these with different procedures.
4. TEFs should not be contained within a specific analytical method, but should be captured in the Standard. TEFs can be captured either directly, with a table which lists the analyte and corresponding TEF, or through reference to TEFs which are evaluated and updated regularly by an internationally recognized procedure.

PARTICIPANTS

NAME	COUNTRY/ ORGANIZATION	EMAIL ADDRESS
Greg Noonan	USA	Gregory.noonan@fda.hhs.gov
Tim Norden	USA	Tim.D.Norden@usda.gov
Marie Maratos	USA	Marie.maratos@fsis.usda.gov
Mr Franz Ulberth	European Union	franz.ulberth@ec.europa.eu
EU Codex Contact point	European Union	codex@ec.europa.eu
Henk A. van der Schee	Netherlands	Henk.van.der.Schee@VWA.NL
Prof. Dr. Peter Koehler	Germany	peter.koehler@tum.de
Dr Andrew P Damant	UK	Andrew.Damant@foodstandards.gsi.gov.uk
Hilde Skår Norli	Norway	nmkl@vetinst.no
Dr. V. T. Gajbhiye	India	head_chem@iari.res.in
Richard Coghlin	Australia	Richard.Coghlan@measurement.gov.au
Karina Budd	Australia	karina.budd@daff.gov.au
Codex Contact Point	Australia	codex.contact@daff.gov.au
Stan Bacler	Canada	stanley.bacler@hc-sc.gc.ca
Jeff Van de Riet	Canada	jeffrey.vanderiet@inspection.gc.ca
Mikko Hofsommer	Int'l Federation of Fruit Juice Producers (IFU)	Info@gfl-berlin.com
Ms. Chanchai Jaengsawang	Thailand	chan48@ymail.com
Ms. Usa Bamrungbhuet	Thailand	usa@acfs.go.th
Ms. Chitrlada Booncharoen	Thailand	chitr@hotmail.com; chitrlada@acfs.go.th
Codex Contact Point	Thailand	Codex@acfs.go.th
Pedro A. Burdaspal	Spain	pburdaspal@msssi.es
Dr. Markus Lipp	USCP	mxl@usp.org
Roger Wood	UK	roger.shirley@btinternet.com
Theodora Klada	Greece	dxyath-additives@gcsl.gr
Astrid Nordbotten	Norway	astrid.nordbotten@mattilsynet.no
Anne Bridges	AACCI	annebridges001@earthlink.net
Paul Wehling	AACCI	Paul.Wehling@genmills.com
Filomena Justino	Angola	fihlomenajustino1@yahoo.com.br
Dr. Stephen Ellison	Eurochem	s.ellison@lgcgroup.com
Lic. Mauricio Gonzalez Zeledon	Costa Rica	gonzalez@senasa.go.cr
Amanda Lasso Cruz	Costa Rica	alasso@meic.go.cr
Dr. A.	Ghana	colecraft_s@hotmail.com
Codex Contact Point	Ghana	codex@gsa.gov.gh
Ms. Dr. Katrin Zur	Germany	katrin.zur@bvl.bund.de
Mr. Tomonori Shiokawa	Japan	codexj@mhlw.go.jp
Dr. Takahiro WATANABE	Japan	tawata@nihs.go.jp
Dr. Takanori Ukena	Japan	takanori_ukena@nm.maff.go.jp;
Codex Contact Point	Japan	codex_maff@nm.maff.go.jp
Malose Daniel Matlala	South Africa	CACPSA@health.gov.za
Darija Vratarić	Croatia	darija.vrataric@mps.hr
Mariam Barsoum Onsy	Egypt	eos_mariam@yahoo.com
Richard Cantril	AOCS	richard.cantrill@aocs.org
Hertha Deutsch	Association Of European Coeliac Societies	hertha.deutsch@utanet.at
Mr Hannu Kiviranta	Finland	hannu.kiviranta@thl.fi
Ms Mirja Hokkanen	Finland	mirja.hokkanen@evira.fi

NAME	COUNTRY/ ORGANIZATION	EMAIL ADDRESS
John Szpylka, Ph.D.	Silliker, Inc.	john.szpylka@silliker.com
Mrs Rosane Maria Franklin Pinto	Brazil	rosane.maria@anvisa.gov.br
Konstantin Eller	Russia	eller@ion.ru
Celso Gonzales Chavesta	Peru	cgonzales@lamolina.edu.pe
Shannon Cooksey	GMA	scooksey@gmaonline.com
Korean Contact Point	Korea	codexkorea@korea.kr
Kang-Bong, Lee	Korea	LKB9703@korea.kr
Eun-Jin, Choi	Korea	cej1@korea.kr
Hae-Jung, Yoon	Korea	hjyoon@korea.kr
Jung-Hyuck, Suh	Korea	rndsuh@korea.kr
Hyung-Soo, Kim	Korea	jungin98@korea.kr
César Omar Gálvez González	Mexico	cgalvez@cofepris.gob.mx
Pamela Suárez Brito	Mexico	psuarez@cofepris.gob.mx

ANNEX 1: From InterAgency Meeting Report (IAM 24) presented (was presented at 33rd Session of CCMAS (CRD 2)

Inter-Agency Meeting Discussion Paper: The CODEX Criteria Approach applied to Operationally Defined ("Empirical") Methods

1. Introduction

CODEX permits the use of the 'criteria approach' for the acceptability of test methods. This approach is advantageous because instead of requiring specific endorsement of a detailed procedure, it permits the use of any method that can be shown to meet stated performance criteria. This in turn provides flexibility for laboratories, allowing innovation and efficiency gains to be implemented quickly.

The Criteria Approach currently requires information on Applicability, Minimum Applicable Range, Limit of Detection and Quantitation, Precision (with criteria for reproducibility relative standard deviation), Recovery and Trueness. All of these performance characteristics can be assessed where a method specifies an analyte or other measured property that can be defined independently of the method of determination (for example, the mass fraction of a given trace metal in a foodstuff). However, in the case of properties that can only be defined in terms of a particular procedure (sometimes called 'operationally defined' or 'empirical') it becomes problematic to assess some performance characteristics - particularly Recovery and Trueness, and sometimes detection capability (LoD, LoQ). This follows from the difficulty of obtaining independent reference values against which to test the performance of the method.

This discussion paper considers the extent to which the CODEX criteria approach is applicable in the case of operationally defined.

2. Terminology

Some terminology will be important for this discussion paper. Most terminology follows established CODEX usage; in particular, familiarity with the definitions of types of methods of analysis is assumed. The following terms additionally appear in this document and merit some explanation.

Operationally defined: Defined by reference to a particular documented procedure. The phrase is most commonly applied to methods of analysis, as in 'operationally defined method', but it will be seen that it can also be usefully applied to measurands.

Measurand: Quantity subjected to measurement. Usually, for the purposes of this document, this will refer to the quantities such as the concentration or mass fraction of an analyte, but also encompasses other properties determined by measurement (Howard Mould Count and pH are not, strictly, analyte concentrations but can be described as 'measu rands').

Empirical method: Alternative term for 'operationally defined method'.

Rational method: A method for which the measurand can be defined without reference to a particular procedure for its determination. Examples include: concentration of aspartame in soft drinks, cadmium mass fraction, aflatoxin B1 content (as mass fraction).

3. Operationally defined measurands

Although the phrase 'operationally defined' is often applied to methods of measurement, as is the related term 'empirical method', it is sometimes useful to draw a distinction between method and measurand. Consider, for example, the determination of crude fibre content. Crude fibre is measured by application of a defined procedure involving a series of digestions under prescribed conditions, followed by an ashing step. The result is typically calculated as the ratio of the mass lost in the ashing step to the mass of the test portion. Because the result depends on the digestion and ashing conditions, the method of determination must be standardised to achieve consistent results across different laboratories. This standardisation, though typically chosen to represent the effects of the human digestive system and decided by expert consensus, is essentially an arbitrary choice of the particular conditions. This is typical of a Type I method. Notice, however, that although the procedure has been carefully defined, the effect of doing so is to define not only the method of measurement, but the measurand. In effect, the measurand 'crude fibre content' is defined as 'the result of applying the defined test method for crude fibre'.

This distinction is important when the possibility of alternative methods is considered. For example, one could conceive of an infrared spectroscopy method being used in routine monitoring or screening of crude fibre content, after calibration against crude fibre values on reference materials established using the defining method. This use of an alternative method is, in effect, the estimation of the operationally defined measurand using a very different method of measurement to the defining method.

The explicit recognition that such a measurand can be estimated by other measurement procedures subject to demonstration of performance is central to the idea of applying the Criteria Approach to potential alternative methods to

Type I methods of analysis. It may therefore be useful to include the distinction in the CODEX guidelines to clarify whether and where it may be permissible to adopt an alternative method of analysis to a Type I method.

4. The Criteria Approach applied to Codex Type I methods 4.1 Codex Type I methods

Codex Type I methods are methods of analysis which 'determine a value that can only be arrived at in terms of the method per se and serve by definition as the only method of establishing the accepted value of the item measured'.

Four observations follow from this description:

i) The description restricts the term 'Type I method' to those measurands that can NOT be defined independently of a test method; that is, it can only refer to operationally defined measurands.

ii) Type I methods, by virtue of their designation as a defining method, always deliver an unbiased estimate of the value of the measurand when correctly used.

iii) It is not possible for an alternative method of measurement to establish a reference value against which a Type I method can be tested

iv) It **IS** possible, at least in principle, for a Type I method to provide reference values against which possible alternative methods can be tested or calibrated.

4.2 Applicability of Trueness criteria to Type I methods

From observation i), it follows that it is not useful to consider the trueness of Type I methods in the sense of closeness of the expected mean value delivered by the method to an independent true value, because the expected mean value of a type I method applied to a test material will, by definition, be the true value.

The Trueness provisions of the Criteria Approach therefore cannot be applied to Type I methods.

It is, however, possible to apply the Criteria Approach to a possible alternative method intended to estimate the Type I value, as trueness or apparent recovery can be established by comparison of results from the Type I method and the proposed alternative. It is, however, likely to be hard to make this assessment for all possible test sample types and levels; it follows that it may be necessary to restrict the scope of application of proposed alternative methods to the range of sample types considered in the evaluation.

4.3 Applicability of Precision criteria to Type I methods

A Type I method that delivers very poor precision is unlikely to be useful in a regulatory context. It may therefore be useful to consider specifying minimum performance criteria for precision when adopting Type I methods if this is not already in place.

Where a reference method is to be used for calibration of routine methods, it is usually necessary to require substantially better performance than for the routine methods it is intended to calibrate (a typical guideline is that a calibration method should achieve uncertainties a factor of three lower than the method it is intended to calibrate). This may indicate a need to establish different performance criteria - whether under the Criteria Approach or not - for Type I methods intended to be used for calibration.

4.4 Applicability of detection capability criteria to Type I methods

Since type I methods define the true value, there is no clear way of establishing that a test material has an identically zero response to the method. This makes it difficult in principle to establish a critical value (the value above which a test material is declared to have nonzero true value). Since the LoD depends directly on the critical value, it might be thought problematic to establish capability of detection.

Fortunately this is rarely a problem in practice. In general, if a proposed Type I method is achieving precision which is, within the range of interest for regulation, sufficient to distinguish a test material from a zero response, the method is likely to be fit for use. Given this information, a strictly determined LoD is unlikely to be necessary.

It is, as above, important that where a proposed alternative to a Type I method is considered, it is tested against materials that cover the full range of intended use for the alternative method. Such methods are, however, necessarily tested against values provided by the Type I method itself and values at the low end of the range are likely to be available.

A potential consideration for the application of the Criteria Approach to Alternative methods is that where the Type I method itself is not applicable near zero, it is not sensible to require LoD determination for a proposed alternative. It should be sufficient to demonstrate acceptable precision at the low end of the operating range f the Type I method.

4.5 Trueness for individual laboratory implementations of Type I methods

It is always useful to consider whether a particular laboratory's implementation of a type I method is unbiased. This is most commonly tested through Proficiency Testing or by use of reference materials certified by the same Type I

method. To assure adequate performance, PT providers may find it useful to consider the numerical values of the Criteria Approach's trueness criteria when setting criteria for scoring PT results, and in particular when reporting on cumulative evidence of bias.

5. Applicability of the Criteria Approach to other Operationally Defined methods

The provisions of the Criteria Approach already apply to Codex Type II ('Reference') and Type III ('Alternative') methods. Type II methods are, by definition, reference methods for situations where Type I methods do not apply - that is, in the case where the method (and hence measurand) can be defined without reference to a measurement procedure. In such cases it is usually possible to test trueness using materials prepared gravimetrically or (increasingly) certified using test methods of very small uncertainty based on unique facilities at National Measurement Institutes. The provisions of the Criteria Approach are therefore expected to apply, subject to the possibility of requiring better performance for calibration or referee methods.

For Type III ('Alternative') methods, it can reasonably be expected that trueness and, if necessary LoD, can be established by comparison with values provided by a Type I or Type II method. Again, the Criteria Approach should apply without change.

6. Conclusions

For Type I methods, the Criteria Approach cannot apply in full. It may, however, be useful to set additional performance criteria, particularly for precision, for establishing Type I methods when the intended use is for calibration.

It may be useful to note explicitly in Codex guidance that Type I methods define a measurand that could in principle be estimated using alternative methods of measurement, subject to demonstration of adequate performance as defined by the Criteria Approach.

No immediate adjustment to the Criteria Approach seems necessary in the case of Type II methods proposed as alternatives to defined Type I or reference methods.

This discussion paper was prepared for the IAM by S Ellison (Eurachem representative to the IAM)

ANNEX 2: From Norwegian Response to CL 2013/16-FFP, 33rd Session CCFFP

Annex 3 Why Horwitz/Thompson equation is not valid for multi component methods[4] 1

The total toxicity is the sum of the concentration of the analogues multiplied with the respective TEF.

At the minimum applicable level, mAL, this is expressed as ∑mAL·TEF

In the Codex Procedural Manual, the criterion for the precision is given as the relative standard deviation. The relative standard deviation, RSD, is expressed as the following:

$$RSD(\%) = \frac{s}{x} \cdot 100\% \Leftrightarrow s = \frac{RSD \cdot x}{100} \qquad (1)$$

where *s* is the standard deviation and *x* is the concentration (here: x= mAL·TEF).

The standard deviation for the total toxicity would be the combined uncertainty of the standard deviation of the analogues, i.e. the sum of the variances of the standard deviation, s^2, of the analogues of interest.

$$s_{total} = \sqrt{\sum s_i^{\,2}} = \sqrt{(\frac{RSD_i}{100} \cdot mAL_i \cdot TEF_i)^2} \qquad (2)$$

The relative standard deviation of the total toxicity, RSD_{total}, at minimum applicable level (mAL), is the square root of the sum of the variances of the individual compounds divided by the concentration:

$$RSD_{total}(\%) = \frac{\sqrt{\sum (\frac{RSD_{Ri}}{100} \cdot mAL_i \cdot TEF_i)^2}}{\sum (mAL_i \cdot TEF_i)} \cdot 100\% = \frac{\sqrt{\sum (RSD_{Ri} \cdot mAL_i \cdot TEF_i)^2}}{\sum (mAL_i \cdot TEF_i)} \qquad (3)$$

Depending on the TEF values, the mALs for PSP are 0.05, 0.1 and 0.5 mg/kg, respectively, (see Annex 2) (corresponding to maximum level of 0.14, 0.25 and 1 mg/kg in the Codex criteria, using the Excel Spreadsheet at the homepage of NMKL). For these levels the RSD will vary from 32 - 44%. The numbers used in the calculations are given in the table below.

[4] This Annex is taken directly from the Norwegian response to CL 2013/16-FFP which is included in Codex document CX/FFP 14/33/5.

Table 3.1 The TEF, mAL and RSD of the PSP toxins, and the combination thereof for the use in the estimation of the relative standard deviation of the total toxicity, RSD_{total}

PSP TOXIN	**TEF**	**mAL (mg/kg)**	**TEF· mAL (mg/kg eq.)**	**RSD(%)**	**$(mAL \cdot TEF \cdot RSD)^2$ $(mg/kg\ eq\ \%)^2$**
STX	1	0.05	0.05	44	4.84
GTX1	1	0.05	0.05	44	4.84
GTX2	0.4	0.1	0.04	39	2.43
GTX3	0.6	0.1	0.06	39	5.48
GTX4	0.7	0.1	0.07	39	7.45
GTX5 (B1)	0.1	0.5	0.05	32	2.56
GTX6 (B2)	0.1	0.5	0.05	32	2.56
dcGTX2	0.2	0.1	0.02	39	0.608
dcGTX3	0.4	0.1	0.04	39	2.43
C1 (epi-GTX8)	0.006	0.5	0.003	32	0.0092
C2 (GTX 8)	0.1	0.5	0.05	32	2.56
C3	0.01	0.5	0.005	32	0.026
C4	0.1	0.5	0.05	32	2.56
NEO	1	0.05	0.05	44	4.84
dcSTX	0.6	0.1	0.06	39	5.48
dcNEO (GTX 7)	0.4	0.1	0.04	39	2.43
11-hydroxy-STX	0.3	0.1	0.03	39	1.37
Sum	**7**		**0.718**		**52.48**

Using formula (3) to estimate the relative standard deviation of the total toxicity, the following is obtained:

$$RSD_{total}(\%) = \frac{\sqrt{4.84 + 4.84 + 2.43 + \ldots + 1.37}}{0.72} = 10\%$$

A relative standard deviation for the total toxicity of 10% is very tight. If only the five first analogues of the PSP toxins in table 3.1 were present (STX, GTX1, GTX2, GTX3 and GTX4), the RSD_{total},would be 19%. If only STX was present, the RSD_{total}, would be 44% . And when having only one analogue, this would be according to the numeric value obtained by using the Horwitz/Thompson equation.

The ML for the total toxicity for the PSP is 0.8 mg/kg STX·diHCl eq. For a single component with concentration of 0.8 mg/kg, the predicted RSD_R is 33%. If this should be the requirement for the total toxicity, and all analogues were present, the RSD of each analogue could be above 100%, which is not satisfactory.

The RSD of the total becomes smaller when the number of components increases. The more components, the narrower the RSD become. This can easy be illustrated if assuming that for *n* analogues the TEF=1, mAL=1 and RSD=44%

$$RSD_{total} = \frac{\sqrt{\sum (RSD_{Ri} \cdot mAL_i \cdot TEF_i)^2}}{\sum (mAL_i \cdot TEF_i)} = \frac{\sqrt{\sum (44_1 \cdot 1_1 \cdot 1_1)^2 + (44_2 \cdot 1_2 \cdot 1_2)^2 + \ldots + (44_n \cdot 1_n 1_n)^2}}{\sum (1_1 \cdot 1_1) + \ldots + (1_n \cdot 1_n)}$$

$$= \frac{\sqrt{44^2 \cdot n}}{n} = \frac{44\sqrt{n}}{n}$$

When n increases, the RSD_{total} becomes smaller.

This shows that Horwitz /Thompson equation cannot be applied to multi component results, nor does it make sense to set criteria for the precision for a sum of components, based on the equation above, as the precision will be smaller though more analogues present.