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 - E-mail: codex@fao.org - www.codexalimentarius.org

Agenda Item 4 **CX/MAS 16/37/4**

January 2016

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING

37th Session

Budapest, Hungary, 22 - 26 February 2016

DEVELOPMENT OF PROCEDURES/GUIDELINES FOR DETERMINING EQUIVALENCY TO TYPE I METHODS

(Prepared by the United States of America)

BACKGROUND

1. At its 35th session, the Codex Committee on Methods of Analysis and Sampling (CCMAS35) considered the "*Discussion Paper on Considering Procedures for Establishing Criteria for Multi-analyte and Type I Methods*".

2. As part of those discussions CCMAS35 agreed that numerical criteria for Type I methods should not be established, however that is might be useful to consider and discuss procedures for establishing equivalency to Type I methods (Para 59 of REP14/MAS).

3. Based on that decision an electronic working group, chaired by the United States of America and working in English was established to develop a discussion paper which would consider different procedures/guidelines for determining equivalency to Type I methods (Para 61 of REP14/MAS).

4. A discussion paper prepared by the United States of America was presented for comment at the CCMAS36.

5. There was general agreement by the Committee that work should continue, but with caution, since equivalency could have unintended consequences. A number of general recommendations were suggested, including:

- Clearly define the concept of equivalent methods and how such equivalency would apply to Type I–IV methods.
- Clarify the role an equivalent method would have in dispute resolution.
- Review other procedures or national protocols for establishing equivalency.

6. Based on the discussion by the Committee, an eWG chaired by the United States of America and working in English, was continued. However, due to other duties, the United States of America was not able to complete a draft discussion paper in time to allow for discussion among the participants of the eWG. Therefore this discussion paper is presented for further discussion to proceed during CCMAS37.

SCOPE

7. The purpose of this paper is to recommend a statistical approach for establishing equivalence to an existing Type I method, outline recommendations for acquiring the analytical results with each method and identify further topics for consideration, discussion and development.

8. These procedures are being developed to enable the establishment of equivalency between a Type I method and any other Type (I-IV) method. Establishing equivalency to a Type I method, does not impart any status to the equivalent method, i.e. the new method does not become "endorsed", by way of equivalency. *[While the equivalent method may be used for testing and control purposes, in the event of a dispute the Codex endorsed Type I method would remain as the defining method.]*

9. While the procedures are primarily intended for establishing equivalency to Type I methods, they will also be applicable to establishing equivalency between any 2 methods, regardless of the Type (e.g. Type II to

III). However, because there are provisions for establishing Numerical Criteria with respect to Type II-IV methods, establishing equivalency between such methods may not be advantageous.

INTRODUCTION

10. The Codex Procedural Manual defines a Type I 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.*". Based on this definition, a Type I method that has been endorsed by Codex is expected to produce the true value for the applicable measurand and would serve as the benchmark for determining the trueness of any alternative to the Type I method.

11. General criteria for Codex methods of analysis have been specified and include the requirement for an inter-laboratory or single laboratory validation to confirm that the method is 'fit for purpose'. This means that a rigorous process has been applied to evaluate the performance characteristics of the method using appropriate samples that define the scope of the method, preferably in multiple laboratories. These same principles should apply to any method that is being considered as equivalent to a Type I method. However, in addition to meeting all of the general criteria as specified in the Procedural Manual, this "alternative" method should also produce results for each sample that are equivalent to the existing Type I method. The question of equivalence is not only if the alternative method is fit for purpose; but also, does the proposed alternative method produce results that are equivalent to the endorsed Type I method?

12. In order to evaluate the equivalence of a proposed alternative to a Type I method, it will be necessary to run one or more sets of samples using both methods and then to compare the results. For this reason, a prerequisite is that the two methods produce results for the same measurand in the same units. Samples run by each method should be homogeneous and cover each representative matrix and concentration range defined in the scope of the alternative method. One sample set should be run for each matrix and concentration level that is necessary to demonstrate equivalency for the new method scope. It is necessary that each and every sample set pass an appropriate test for equivalence in order for the alternative method to be considered truly equivalent.

13. Unfortunately, there is little guidance from regulatory agencies or scientific associations such as AOAC or ISO on the exact procedures for establishing the equivalence of analytical methods. In recent years, the pharmaceutical industry has been required to establish procedures for evaluating method equivalence based on regulations for the bioequivalence of orally administered drug products. As a result, the issue of method equivalency has been thoroughly discussed in this context and the discussion of procedural options that follows is taken largely from a review of these papers.

STATISTICAL APPROACH

Two one-sided *t*-Test (TOST)

14. A statistical test called the two one-side t-test (TOST) is being proposed as the recommended statistical test for evaluating equivalency of two methods. TOST is being recommended over a two-sample t-test because it is considered more rigorous. Additionally, problems related to the two-sample *t*-test are solved by using TOST due to the specification of the acceptance criterion, θ. Specifically, if a method has poor precision or a small number of samples are used when evaluating equivalency, the two-sample t-test can erroneously identify the methods as equivalent. Alternatively, the two-sample t-test can lead to the conclusion of a statistically significant difference even when the difference is of no practical importance.

15. The TOST begins with the null hypothesis; that the two mean values for the methods are not equivalent. A positive test for significance then results in demonstrating, at a specified confidence level, that the two data sets are equivalent. The TOST test requires the specification of a parameter called the acceptance criterion, $\pm\theta$, which represents the smallest difference in mean values for the two methods that is deemed as practically important. The null hypothesis, H_0, and the alternate, H_a, are described in terms of the difference in means, $\mu_1 - \mu_2$, and θ by the following

$$H_0: \mu_1 - \mu_2 \leq \theta_L \; or \; \mu_1 - \mu_2 \geq \theta_U$$

$$H_a: \theta_L < \mu_1 - \mu_2 < \theta_U$$

16. The alternative hypothesis is proven at a specified level of confidence when the true difference in means between the methods is within the boundaries specified by $\pm\theta$.

17. TOST is carried out by first selecting/calculating a θ, which is appropriate for the methods and has practical relevance.

$$\theta = \delta + s^*[t_{(1-\alpha,2n-2)} + t_{(1-\beta,2n-2)}]\sqrt{\frac{2}{n}} \qquad (1)$$

Where δ is the absolute value of the true difference between the groups' mean values. Often set to zero in the most conservative approaches. To ensure that the true measurement precision is represented, the upper confidence limit, s*, is used in place of s and is calculated by

$$s^* = s\sqrt{\frac{n-1}{\chi^2_{(\gamma,n-1)}}} \qquad (2)$$

In which $\chi^2_{(\gamma,n-1)}$ is the (100γ)th percentile of a distribution with n-1 degrees of freedom.

Once θ is specified, the confidence interval (CI) for the difference in means at a specific level of confidence (usually 95%) is calculated by

$$CI = \bar{x}_1 - \bar{x}_2 \pm t_{90,(n_1+n_2-2)} \cdot \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad (3)$$

Where $\bar{x}_1$ and $\bar{x}_2$ are the mean values from each sample set for each method, $t_{90,(n_1+n_2-2)}$ is the t-value at 90% confidence with n_1+n_2-2 degrees of freedom, s_p is the estimate for the pooled standard deviation for the sample sets and $n_1$and n_2 are the number of samples run in each set. The value for $t_{90,(n_1+n_2-2)}$ can be found in statistical tables or by using the TINV function in Excel. (Note that the 95% one-sided confidence limit is equivalent to the 90% two-sided confidence limit). If CI is completely contained within the range defined by $\pm\theta$, the methods as defined by this data set are deemed equivalent.

18. One drawback of the TOST, especially in attempting to set general procedures, is the need to specify/calculate acceptance criterion (θ) relevant to each method comparison. It may be difficult to establish a single acceptance criterion for all methods, since it requires the identification of the smallest practically significant difference in the mean values and this may require an individual assessment for each method. While equation 1 can be used to assist the analyst in determining θ, it is possible that leaving the determination of θ to the analysts could lead to incorrect equivalency decisions. What also may be considered a detriment of the TOST is the number of samples necessary to obtain the adequate power or confidence in the conclusion. The number of samples needed is dependent on the ratio of θ to s_p with larger ratios requiring a smaller number of samples. When θ and s_p are similar in value, 18-27 samples are required to obtain sufficient power for the equivalence test.[9]

STATISTICAL RECOMMENDATIONS

19. The TOST should be used to establish equivalency between 2 methods.

20. Set δ to zero unless specific conditions are identified by the analyst to justify a nonzero value.

21. Set α and β to 0.05 representing a 95% confidence limit unless specific conditions or practical considerations justify a change.

22. The acceptance criterion (θ) will need to be determined by the analyst for each method comparison, but a small table containing frequent values for θ could be included in the document to guide in that determination.

ANALYTICAL TESTING RECOMMENDATIONS

23. Prior to performing the statistical comparison, analysis of multiple sample sets that include the appropriate matrix and concentration levels to cover the scope of the methods would be required. It is suggested that at least three concentrations be run for each relevant matrix, but additional samples run across various concentrations would be advantageous. One set should be at the limit of quantitation (LOQ) and a second set should be at the ML (minimum or maximum level), with the third being at the high end of the analytical method. If the method does not have a specified ML, a low, medium and high concentration should be chosen that covers the concentration range of the method.

24. For the TOST procedure, the number of samples to be run in order to demonstrate equivalence varies with the standard deviation and the acceptance criterion, but a conservative estimate of 18-27 samples for each data set would cover situations when θ is similar in magnitude to s_p. The equivalence studies should be performed under reproducibility conditions as much as possible. The studies could be performed in a single laboratory, but it would be desirable to utilize different analysts, materials, and equipment as much as possible.

25. Since TOST assumes that the two methods have equal variance, the test results should be used to confirm this by using the F-test. It would be advantageous to perform this test for equal variance at each concentration (low medium high), in addition to the entire range.

26. Equivalency should be tested for each matrix analyzed because even small differences in matrix may impact the results differently for different methods.

QUESTIONS FOR DISCUSSION

27. This paper recommends a statistical approach for evaluating the equivalency between two methods and outlines a few general recommendations for acquiring the necessary analytical results. While this approach is developed with the intent of showing equivalency between a Type I method and any other method Type (I-IV), it could also be applied to establishing equivalency between any 2 methods, regardless of Type. The paper clearly does not address specific details in implementing this approach or list a stepwise procedure. Additionally, there are a number of questions that should be considered prior to developing the final procedures.

i.) If general procedures for evaluating equivalence are established, where will they reside? In the Codex Procedural Manual or in a Guidance/Information document?

ii.) For methods measuring a composition or characteristic (e.g. moisture content) it would be required that the two methods be equivalent across the entire range of the method. However, for provisions where a maximum limit is established would it be acceptable to establish equivalency around that limit, but not worry about equivalency at some value well above the limit?

iii.) This paper has focused primarily on quantitative methods, but procedures for qualitative may also be useful. Such procedures would have a very different format/approach, so would they be included in a single document or would separate documents be developed for quantitative and qualitative methods?

Bibliography

a) Freund, J. E.; *Modern Elementary Statistics*, 7th ed.; Prentice-Hall: Englewood Cliffs, New Jersey, 1988; p 313.

b) Zhong, J.; Lee, K.; Tsong, Y.; *J. Biopharm. Stat*. **2008**, *18*, 1005-1012.

c) Bland, D. G.; Altman, J. M. *Lancet* **1986,** *8,* 255.

d) Bland, D. G.; Altman, J. M.; *Stat. Methods Med. Res.* **1999,** *8,* 137-160.

e) Dewe, W.; *J. Chrom. B* **2009**, *877*, 2208-2213.

f) Hartmann, C.; Smeyers-Verbeke, J.; Penninckx, W.; Vander Heyden, Y.; Vankeerberghen, P.; Massart, D. L. *Anal. Chem*. **1995**, *67*, 4491-4499.

g) Schuirmann, D.; *J. Pharmaco. Biopharm*. **1987**, *15*, 657-680.

h) Chatfield, M. J.; Borman, P. J. *Anal. Chem.* **2009**, *81*, 9841-9848.

i) Borman, P. J.; Chatfield, M. J.; Damjanov, I.; Jackson, P. *Anal. Chem.* **2009**, *81*, 9849-9857.

j) Chambers, D.; Kelly, G.; Limentani, G.; Lister, A.; Lung, K. R.; Warner, E. *Pharm. Tech*. **2005**, *29*, 64-80.

k) Limentani, G. B.; Ringo; M. C., Ye, F.; Bergquist, M. L.; McSorley, E. O. *Anal. Chem.* **2005**, *77*, 221A-226A.