



**PROGRAMME MIXTE FAO/OMS SUR LES NORMES ALIMENTAIRES
COMITÉ DU CODEX SUR LES MÉTHODES D'ANALYSE ET D'ÉCHANTILLONNAGE**

Trente-sixième session

Budapest, Hongrie, 23-27 février 2015

**DOCUMENT DE TRAVAIL SUR L'ÉLABORATION DES PROCÉDURES/DIRECTIVES POUR
DÉTERMINER L'ÉQUIVALENCE AUX MÉTHODES DE TYPE I**

(Soumis par les États-Unis d'Amérique)

RAPPEL DES FAITS

1. À sa trente-cinquième session, le Comité du Codex sur les méthodes d'analyse et d'échantillonnage (CCMAS) a examiné le «*Document de travail sur l'examen des procédures d'établissement des critères pour des méthodes multi-analytes et des méthodes de Type I*».
2. Au cours des débats, le CCMAS est convenu qu'il n'était pas nécessaire d'établir de critères numériques pour les méthodes de type I, mais qu'il pouvait par contre être utile d'examiner des procédures d'établissement de l'équivalence aux méthodes de Type I (paragraphe 59 du document REP14/MAS).
3. Compte tenu de cette décision, un groupe de travail électronique, présidé par les États-Unis d'Amérique et travaillant en anglais, a été constitué et chargé d'élaborer un document de travail sur les différentes procédures et directives permettant de déterminer l'équivalence aux méthodes de type I (paragraphe 61 du document REP14/MAS).
4. Le groupe de travail électronique auquel ont participé 20 délégués, a suscité un intérêt considérable. Toutefois, en raison d'autres obligations, les États-Unis d'Amérique n'ont pas été en mesure d'achever en temps voulu un projet de document qui aurait permis aux membres du groupe de travail électronique d'en débattre. Le présent document constitue donc une première version qui est soumise pour observation et examen au CCMAS, à sa trente-sixième session.

INTRODUCTION

5. Selon la définition du Manuel de procédure, une méthode de type I est «*une méthode qui définit une valeur qu'il n'est possible d'obtenir qu'aux termes de la méthode per se et qui est, par définition, la seule utilisée pour établir la valeur acceptée de l'élément mesuré*». (Vingt et unième édition, 2013, page 69). Compte tenu de cette définition, une méthode de type I qui a été approuvée par le Codex est supposée produire la valeur vraie du mesurande applicable et doit servir de référence pour déterminer la justesse de toute autre méthode de remplacement de la méthode de type I.
6. L'objet du présent document est d'examiner les procédures permettant d'établir l'équivalence à une méthode de type I.
7. Des critères généraux ont été spécifiés pour les méthodes d'analyse du Codex et comportent notamment la validation entre laboratoires ou par un laboratoire unique afin de confirmer que la méthode est apte au but poursuivi. Cela signifie qu'un processus rigoureux a été appliqué, de préférence dans différents laboratoires, pour évaluer les caractéristiques de performance de la méthode en utilisant des échantillons appropriés qui définissent le champ d'application de la méthode. Ces mêmes principes doivent s'appliquer à toute méthode qui est considérée comme équivalente à une méthode de type I. Toutefois, cette méthode de remplacement doit non seulement répondre à tous les critères généraux énoncés dans le Manuel de procédure, mais aussi produire des résultats pour chaque échantillon qui sont équivalents à ceux de la méthode de type I existante. La question qui se pose au sujet de l'équivalence ce n'est pas seulement de savoir si la méthode de remplacement est apte au but poursuivi mais si elle produit des résultats qui sont équivalents à ceux de la méthode de type I approuvée.
8. Afin d'évaluer l'équivalence d'une méthode proposée en remplacement d'une méthode de type I, il faut analyser une ou plusieurs séries d'échantillons à l'aide des deux méthodes et comparer ensuite les résultats

obtenus. C'est pourquoi les deux méthodes doivent impérativement produire des résultats pour le même mesurande dans les mêmes unités. Les échantillons analysés par chaque méthode doivent être homogènes et couvrir toutes les matrices et toutes les fourchettes de concentration représentatives définies dans le champ d'application de la méthode de remplacement. Une série d'échantillons doit être analysée pour chaque matrice et pour chaque fourchette de concentration nécessaires pour établir l'équivalence du champ d'application de la nouvelle méthode. Chaque série d'échantillons doit impérativement être soumise à un test d'équivalence approprié pour que la méthode de remplacement puisse être considérée comme vraiment équivalente.

9. Il est cependant regrettable qu'il n'existe que peu d'indications de la part des organismes de réglementation ou des associations scientifiques comme AOAC ou ISO sur les procédures exactes permettant d'établir l'équivalence des méthodes d'analyse. Ces dernières années, il a été demandé à l'industrie pharmaceutique d'établir des procédures d'évaluation de l'équivalence des méthodes en fonction des réglementations concernant la bioéquivalence des médicaments administrés par voie orale. La question de l'équivalence des méthodes a donc fait l'objet de débats approfondis et l'examen des procédures possibles qui est présenté ci-après est largement repris des documents publiés à ce sujet.

APPROCHES STATISTIQUES

Option 1: Test t à deux échantillons

10. Une solution pour évaluer l'équivalence d'une méthode consiste à comparer les valeurs moyennes des ensembles de données pour chaque méthode en déterminant la signification statistique à l'aide du test t à deux échantillons^{1, 11}. Il s'agit d'un test statistique dans lequel l'hypothèse nulle ou supposition initiale est que les moyennes des méthodes sont égales. Le test vise à infirmer cette hypothèse, et s'il y réussit, prouve à un niveau de confiance déterminé (en général 95 pour cent) que les moyennes sont significativement différentes. Si le test échoue, montrant qu'il n'existe pas de différence statistique des moyennes entre les ensembles de données, les moyennes sont supposées égales par défaut. Le test suppose que les deux ensembles de données contiennent des résultats qui sont distribués normalement, que les écarts-types sont semblables et que la distribution t de Student décrit la variation aléatoire pour des tailles d'échantillon inférieures à $n = 30$. Pour effectuer ce test, la valeur de la statistique t est calculée par

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (1)$$

où \bar{x}_1 et \bar{x}_2 sont les valeurs moyennes de chaque série d'échantillons pour chaque méthode, s_p est l'estimation de l'écart-type cumulé pour chaque série d'échantillons, et n_1 et n_2 sont les nombres d'échantillons analysés dans chaque série. s_p pour les séries de mesures répétées est égal à

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (2)$$

où s_1 et s_2 sont les écarts types de chaque série d'échantillons. La valeur t tirée de l'équation (1) est calculée et comparée à la valeur critique t pour un niveau de confiance déterminé (en général 95 pour cent). On peut trouver cette valeur dans des tableaux statistiques ou en utilisant la fonction TINV dans Microsoft Excel. Si la valeur absolue de la valeur t calculée est supérieure ou égale à la valeur t critique, les deux ensembles de données sont déclarés comme étant statistiquement différents.

11. Ce même test peut être effectué selon une seconde méthode qui consiste à calculer l'intervalle de confiance (CI) pour la différence entre les deux valeurs moyennes en utilisant la formule

$$CI = \bar{x}_1 - \bar{x}_2 \pm t_{95, (n_1 + n_2 - 2)} \cdot \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (3)$$

où \bar{x}_1 et \bar{x}_2 sont les valeurs moyennes de chaque série d'échantillons pour chaque méthode, $t_{95, (n_1 + n_2 - 2)}$ est la valeur t à un niveau de confiance de 95 pour cent avec $n_1 + n_2 - 2$ degrés de liberté, s_p est l'estimation de l'écart-type cumulé pour les séries d'échantillons donnée par l'équation (2), et n_1 et n_2 sont les nombres d'échantillons analysés dans chaque série. On peut trouver la valeur de $t_{95, (n_1 + n_2 - 2)}$ dans des tableaux statistiques ou en utilisant la fonction TINV dans Microsoft Excel. Si CI ne comprend pas 0, les moyennes sont déclarées comme étant significativement différentes.

12. On a fait remarquer que le test t à deux échantillons pouvait ne pas être approprié pour établir l'équivalence de deux méthodes, étant donné qu'il était conçu pour prouver que deux ensembles de données ont des moyennes différentes, et non qu'ils ont des moyennes équivalentes.^{6, 8, 9, 10, 11.} L'hypothèse nulle ou par défaut étant que les moyennes sont égales, la méthode ne peut fournir d'éléments probants que si les deux moyennes sont inégales. Lorsqu'il n'y a pas de preuves suffisantes indiquant une différence significative entre les moyennes, l'hypothèse nulle est acceptée par défaut ce qui ne prouve pas pour autant que les moyennes sont réellement égales. L'application du test t à deux échantillons pour évaluer l'équivalence des méthodes soulève deux problèmes particuliers. Premièrement, lorsque les écarts types des deux ensembles de données sont relativement élevés et que le nombre des mesures est relativement bas, le test t peut indiquer une différence peu significative alors qu'avec des mesures supplémentaires il aurait montré que les moyennes étaient en réalité statistiquement différentes. Deuxièmement, lorsque les écarts types sont relativement faibles, le test peut indiquer une différence statistique alors que la différence absolue est pratiquement peu importante.

Option 2: Deux tests t unilatéraux (TOST)

13. La procédure dite des deux tests unilatéraux a été proposée comme solution de rechange recommandée au test t à deux échantillons. Elle commence par l'hypothèse nulle contraire, à savoir que les deux valeurs moyennes pour les méthodes ne sont pas équivalentes^{6, 7, 11.} Un test de signification positif permet de montrer que, à un certain niveau de confiance, les deux ensembles de données sont équivalents. Un paramètre intitulé critère d'acceptation, $\pm\theta$, doit être spécifié pour cette procédure; il représente la plus petite différence entre les valeurs moyennes des deux méthodes qui est jugée comme étant pratiquement importante. L'hypothèse nulle, H_0 , et l'hypothèse alternative, H_a , sont décrites en termes de différence des moyennes, $\mu_1 - \mu_2$, et θ par

$$H_0: \mu_1 - \mu_2 \leq \theta_L \text{ or } \mu_1 - \mu_2 \geq \theta_U$$

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

14. L'hypothèse alternative est démontrée à un niveau de confiance déterminé lorsque la différence réelle entre les moyennes obtenues avec les méthodes se situe dans les limites spécifiées par $\pm\theta$.

15. La procédure consiste d'abord à préciser θ et ensuite à calculer l'intervalle de confiance (CI) pour la différence des moyennes à un niveau de confiance déterminé (en général 95 pour cent) comme suit

$$CI = \bar{x}_1 - \bar{x}_2 \pm t_{90, (n_1+n_2-2)} \cdot \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (4)$$

où \bar{x}_1 et \bar{x}_2 sont les valeurs moyennes de chaque série d'échantillons pour chaque méthode, $t_{90, (n_1+n_2-2)}$ est la valeur t à un niveau de confiance de 90 pour cent avec $n_1 + n_2 - 2$ degrés de liberté, s_p est l'estimation de l'écart-type cumulé des séries d'échantillons donnée par l'équation (2) et n_1 et n_2 sont les nombres d'échantillons analysés dans chaque série. On peut trouver la valeur de $t_{90, (n_1+n_2-2)}$ dans des tableaux statistiques ou en utilisant la fonction TINV dans Excel. (Noter que la limite de confiance unilatérale de 95 pour cent est équivalente à la limite de confiance bilatérale de 90 pour cent). Si CI entre entièrement dans la fourchette définie par $\pm\theta$, les méthodes définies par cet ensemble de donnée sont considérées comme équivalentes.

16. La procédure des deux tests t unilatéraux permet de résoudre les deux problèmes posés par le test t à deux échantillons grâce à la spécification du critère d'acceptation, θ . Le problème soulevé par l'identification d'une différence statistique des moyennes qui n'est pas importante dans la pratique est résolu puisque la spécification de la fourchette ($\pm\theta$) permet un biais statistiquement significatif tant qu'il est suffisamment petit pour s'inscrire dans la fourchette. Le cas où les valeurs s de la méthode sont relativement importantes est également pris en compte, étant donné que l'équivalence de la méthode n'est établie que si l'intervalle de confiance CI calculé pour les deux méthodes est compris entièrement dans la fourchette définie par $\pm\theta$.

17. La procédure des deux tests unilatéraux comporte aussi la spécification du critère d'acceptation, ce qui peut être perçu comme un inconvénient. L'établissement d'un critère d'acceptation unique pour toutes les méthodes peut être difficile, étant donné qu'il faut pour cela déterminer la plus petite différence significative dans la pratique des valeurs moyennes et probablement évaluer chaque méthode.

18. Un nombre suffisant d'échantillons doit être analysé pour que la conclusion d'un test de signification statistique ait la puissance ou la confiance voulue. Ce nombre est fonction du rapport entre θ et s_p , les rapports les plus élevés nécessitant un nombre d'échantillons plus petit. Lorsque θ et s_p ont des valeurs

semblables, 18 à 27 échantillons sont nécessaires pour obtenir une puissance du test d'équivalence suffisante⁹.

Option 3: Limites de l'accord

19. La troisième option pour évaluer l'équivalence est une approche fondée sur le calcul des limites de l'accord (*LOA*) autour de la méthode de type I approuvée^{2, 3, 4, 5}. Pour chaque ensemble de données analysé, les limites de l'accord (*LOA*) calculées représentent l'intervalle autour de la moyenne dans lequel 95 pour cent des résultats sont supposés se situer. Comme avec les autres procédures, cela suppose une distribution normale et un nombre suffisant d'échantillons afin d'obtenir des estimations raisonnablement bonnes des valeurs moyennes et de l'écart type pour chaque méthode.

$$LOA = u \pm z_{95} \cdot \sigma = 1.96 \cdot \sigma$$

où u est la moyenne, z_{95} est le 95e centile de la distribution normale, et σ est l'écart type de l'ensemble de données pour la méthode de type I approuvée.

20. Ces limites de l'accord (*LOA*) vont alors servir de fourchette de référence pour les valeurs obtenues par la méthode de remplacement. Si 95 pour cent de ces valeurs se situent dans les limites de l'accord, la méthode de remplacement est jugée équivalente pour cette série de données particulière. Afin que le niveau de confiance de ce test particulier soit suffisant, il est recommandé d'analyser 20 échantillons au minimum pour chaque ensemble de données étant donné que 19/20 dans cette fourchette constitue 95 pour cent des résultats. La méthode des limites de l'accord permet un biais entre les méthodes, mais seulement s'il est assez faible pour que 95 pour cent des résultats se situent dans les limites de l'accord (*LOA*) définies par la méthode de type I approuvée.

RÉSUMÉ

21. L'avantage du test t à deux échantillons est qu'il ne nécessite pas d'autres paramètres comme par exemple un critère d'acceptation. Toutefois, le test t n'est pas le test statistique approprié lorsqu'il s'agit d'établir l'équivalence. Il peut notamment amener à conclure que les méthodes sont statistiquement différentes alors que la différence absolue n'est pas importante dans la pratique. Il peut aussi mener à la conclusion que deux méthodes sont équivalentes alors que le nombre d'échantillons analysés est insuffisant pour établir la différence statistique réelle.

22. La procédure des deux tests unilatéraux est probablement la plus rigoureuse pour ce qui concerne l'équivalence statistique, mais elle nécessite la spécification d'une limite d'acceptation pratique qui représente le biais le plus petit entre les méthodes pouvant être acceptées. Dans la pratique, il est probablement difficile d'établir un ensemble de critères d'acceptation pour toutes les méthodes Codex, étant donné qu'elles ont toutes une application spécifique. Une solution consisterait à utiliser des facteurs de récupération figurant dans le Manuel de procédure du Codex à la section «Directives pour l'établissement de valeurs numériques pour le critère» en utilisant la concentration appropriée. Par exemple, le facteur de récupération spécifié de 95 à 105 pour cent pour des concentrations de $\geq 0,1$ suggère un critère d'acceptation de ± 5 pour cent. Cependant, de nombreuses méthodes n'auront pas un facteur de récupération applicable. Il faudrait donc que le comité du Codex s'occupant de produits qui recommande la méthode décide du biais minimal acceptable au cas par cas.

23. L'approche des limites de l'accord (*LOA*) qui implique de calculer un intervalle de probabilité de 95 pour cent autour de la moyenne de la méthode de type I existante semble être une option possible. Elle permet un biais seulement si la méthode de remplacement produit aussi un écart type plus petit de sorte que 95 pour cent des résultats individuels se situent encore dans la fourchette définie par les limites de l'accord (*LOA*). L'avantage de cette approche est qu'elle est très simple à calculer et qu'elle ne demande pas d'établir des critères d'acceptation. Son inconvénient est qu'elle ne permet pas de biais dans la méthode de remplacement si l'écart type de cette dernière est égal à celui de la méthode approuvée. L'équivalence des méthodes de remplacement peut être difficile à établir à l'aide de cette procédure étant donné les ensembles de données multiples représentant des matrices et des concentrations différentes qui doivent être comparés.

24. Chacune de ces procédures possibles nécessite l'analyse de séries d'échantillons multiples comprenant la matrice et les niveaux de concentration appropriés afin de couvrir le champ d'application de la méthode. Il est recommandé d'analyser au moins deux concentrations pour chaque matrice pertinente. L'une doit être à la LM (limite minimale ou maximale) et l'autre à la limite de quantification. Si aucune LM n'est spécifiée pour la méthode, il faut choisir une concentration faible et une concentration forte qui couvrent la fourchette de concentration de la méthode. Pour ce qui concerne la procédure des deux tests unilatéraux, le nombre d'échantillons à analyser pour établir l'équivalence varie en fonction de l'écart-type et du critère d'acceptation mais, selon une estimation prudente, 18 à 27 échantillons pour chaque ensemble de données

devraient couvrir les cas où θ est du même ordre de grandeur que s_p . Au moins 20 échantillons sont recommandés pour l'approche des limites de l'accord (LOA) afin d'obtenir une estimation de l'écart type raisonnablement satisfaisante. Les études d'équivalence doivent autant que possible être effectuées dans des conditions de reproductibilité. Elles peuvent être réalisées dans un laboratoire unique, mais il est souhaitable que les analystes, les matériaux et le matériel soient si possible différents.

QUESTIONS À EXAMINER

25. Le document décrit de façon générale plusieurs approches statistiques permettant d'établir l'équivalence entre des méthodes, mais il est clair qu'il n'entre pas dans le détail de leur mise en application. S'il est possible de recommander une approche générale, des questions restent en suspens quant à son application à des méthodes spécifiques. Par exemple:

- i. Compte tenu de la discussion et de l'éventail des méthodes, peut-on dans la pratique établir un ensemble de critères d'équivalence pour toutes les méthodes Codex?
- ii. Si de tels critères ou mêmes des procédures générales pour l'évaluation de l'équivalence étaient établis où devraient-ils figurer dans le Codex, dans le Manuel de procédure ou dans un document d'orientation?

Références

1. Freund, J. E.; *Modern Elementary Statistics*, 7^{ème} éd.; Prentice-Hall: Englewood Cliffs, New Jersey, 1988; p 313.
2. Zhong, J.; Lee, K.; Tsong, Y.; *J. Biopharm. Stat.* **2008**, *18*, 1005-1012.
3. Bland, D. G.; Altman, J. M. *Lancet* **1986**, *8*, 255.
4. Bland, D. G.; Altman, J. M.; *Stat. Methods Med. Res.* **1999**, *8*, 137-160.
5. Dewe, W.; *J. Chrom. B* **2009**, *877*, 2208-2213.
6. Hartmann, C.; Smeyers-Verbeke, J.; Penninckx, W.; Vander Heyden, Y.; Vankeerberghen, P.; Massart, D. L. *Anal. Chem.* **1995**, *67*, 4491-4499.
7. Schuirmann, D.; *J. Pharmaco. Biopharm.* **1987**, *15*, 657-680.
8. Chatfield, M. J.; Borman, P. J. *Anal. Chem.* **2009**, *81*, 9841-9848.
9. Borman, P. J.; Chatfield, M. J.; Damjanov, I.; Jackson, P. *Anal. Chem.* **2009**, *81*, 9849-9857.
10. Chambers, D.; Kelly, G.; Limentani, G.; Lister, A.; Lung, K. R.; Warner, E. *Pharm. Tech.* **2005**, *29*, 64-80.
11. Limentani, G. B.; Ringo; M. C., Ye, F.; Bergquist, M. L.; McSorley, E. O. *Anal. Chem.* **2005**, *77*, 221A-226A.