



**PROGRAMA CONJUNTO FAO/OMS SOBRE NORMAS ALIMENTARIAS  
COMITÉ DEL CODEX SOBRE MÉTODOS DE ANÁLISIS Y TOMA DE MUESTRAS**

**36.ª reunión**

**Budapest (Hungria), 23-27 de febrero de 2015**

**DOCUMENTO DE DEBATE SOBRE LA ELABORACIÓN DE PROCEDIMIENTOS O DIRECTRICES  
PARA DETERMINAR LA EQUIVALENCIA CON LOS MÉTODOS DEL TIPO I**

**(Preparado por los Estados Unidos de América)**

**ANTECEDENTES**

1. En su 35.ª reunión, el Comité del Codex sobre Métodos de Análisis y Toma de Muestras (CCMAS) examinó el documento titulado “*Discussion Paper on Considering Procedures for Establishing Criteria for Multi-analyte and Type I Methods*” (Documento de debate sobre la consideración de procedimientos para establecer criterios para métodos de analitos múltiples y métodos de Tipo I).
2. A raíz de tales deliberaciones, el CCMAS acordó que no deberían elaborarse criterios numéricos para los métodos de Tipo I; no obstante, señaló que tal vez sería conveniente estudiar procedimientos para establecer una equivalencia con dichos métodos (REP14/MAS, párr. 59).
3. Basándose en esa decisión, se creó un grupo de trabajo electrónico (GTe), presidido por los Estados Unidos de América, que trabajaría en inglés, con el objetivo de preparar un documento de debate en el que se examinarían diferentes procedimientos o directrices para determinar una equivalencia con los métodos de Tipo I (REP14/MAS, párr. 61).
4. Se manifestó un gran interés en el GTe, que contó con 20 participantes. Sin embargo, debido a sus otras tareas los Estados Unidos de América no lograron finalizar el borrador de documento de debate a tiempo para que los participantes del GTe pudieran examinarlo. Por lo tanto, este documento de debate se presenta como un primer borrador, para recabar observaciones y examinarlo nuevamente en la 36.ª reunión del CCMAS.

**INTRODUCCIÓN**

5. En el Manual de Procedimiento se proporciona la siguiente definición de método de Tipo I: “*método que determina un valor al que puede llegarse sólo mediante la aplicación del método en cuestión y que, por definición, es el único método para establecer el valor aceptado del parámetro medido.*” (21.ª ed. 2013, versión española, p. 68). De acuerdo con esta definición, se espera que el método de Tipo I aprobado por el Codex produzca el valor verdadero del parámetro aplicable y sirva como criterio de referencia para determinar la veracidad de cualquier alternativa al método de Tipo I.
6. La finalidad de este documento es examinar los posibles procedimientos para establecer una equivalencia con un método de Tipo I existente.
7. Se han establecido criterios generales para los métodos de análisis del Codex, entre ellos, el requisito de validación en varios laboratorios, o en uno solo, para confirmar que el método en cuestión es “apto para el uso previsto”. Ello supone la aplicación de un proceso riguroso para evaluar las características de rendimiento del método utilizando muestras adecuadas que definan el alcance del mismo, preferiblemente en varios laboratorios. Estos mismos principios deberían aplicarse a cualquier método considerado equivalente a un método de Tipo I. Sin embargo, además de cumplir todos los criterios generales especificados en el Manual de Procedimiento, el método “alternativo” también debería producir resultados, para cada muestra, equivalentes al método de Tipo I existente. La cuestión de la equivalencia no reside únicamente en determinar si el método alternativo es apto para el uso previsto sino también si el método alternativo propuesto produce resultados equivalentes a los del método de Tipo I aprobado.

8. Con el fin de evaluar la equivalencia de una propuesta alternativa a un método de Tipo I, será necesario analizar uno o más conjuntos de muestras utilizando ambos métodos y, posteriormente, comparar los resultados. Por esta razón, un requisito indispensable es que ambos métodos produzcan resultados para el mismo parámetro en las mismas unidades. Las muestras a las que se aplica cada método deberían ser homogéneas y abarcar todas las matrices y todos los rangos de concentraciones representativos definidos en el alcance del método alternativo. Debería analizarse un conjunto de muestras para cada matriz y rango de concentración necesarios a fin de demostrar la equivalencia respecto al ámbito del nuevo método. Es necesario someter cada conjunto de muestras a una prueba apropiada de equivalencia para poder considerar el método alternativo realmente equivalente.

9. Lamentablemente, se dispone de escasa orientación proporcionada por organismos de reglamentación o asociaciones científicas tales como la Asociación de Químicos Analíticos Oficiales (AOAC) o la Organización Internacional de Normalización (ISO) sobre los procedimientos exactos para establecer la equivalencia de métodos analíticos. En los últimos años se ha solicitado a la industria farmacéutica que establezca procedimientos para evaluar la equivalencia de métodos con arreglo a reglamentos de bioequivalencia de medicamentos administrados por vía oral. Como resultado de ello, la cuestión de la equivalencia de métodos se ha debatido a fondo en este contexto; el análisis de las opciones de procedimiento que figura a continuación se ha tomado en gran medida de un examen de esos documentos.

## ENFOQUES ESTADÍSTICOS

### Opción 1: Ensayo $t$ de dos muestras

10. Una opción para evaluar la equivalencia de métodos es comparar los valores medios de los conjuntos de datos de cada uno mediante la realización de una prueba de significación estadística, utilizando el ensayo  $t$  de dos muestras<sup>1, 11</sup>. Se trata de un contraste de hipótesis estadísticas en que la hipótesis nula o suposición inicial es la igualdad de los promedios de los métodos en cuestión. El ensayo trata de refutar esta hipótesis y, si lo logra, demuestra un determinado nivel de confianza (normalmente del 95 %) de que los promedios son considerablemente diferentes. Si el ensayo no es satisfactorio, refleja que no existe una diferencia estadística en los promedios entre los conjuntos de datos; se supone por defecto que los promedios son iguales. En el ensayo se presume que ambos conjuntos de datos contienen resultados con una distribución normal, que las desviaciones estándar son similares y que la distribución  $t$  de Student describe la variación aleatoria de tamaños de muestras inferiores a  $n = 30$ . Para realizar esta prueba, el valor estadístico  $t$  se determina con arreglo a la siguiente fórmula:

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

donde  $\bar{x}_1$  y  $\bar{x}_2$  son los valores medios de cada conjunto de muestras de cada método;  $s_p$  es la estimación de la desviación estándar combinada para cada conjunto de muestras;  $n_1$  y  $n_2$  son el número de muestras objeto de análisis de cada conjunto.  $s_p$  para los conjuntos de mediciones replicadas se calcula con arreglo a la siguiente fórmula:

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

donde  $s_1$  y  $s_2$  son los valores de la desviación estándar de cada conjunto de muestras. El valor  $t$  se calcula aplicando la ecuación 1) y se compara con el valor crítico  $t$  para un determinado nivel de confianza (normalmente del 95 %). Este valor puede encontrarse en tablas estadísticas o utilizando la función T.INV de Microsoft Excel. Si el valor absoluto del valor calculado  $t$  es mayor o igual al valor crítico  $t$ , se concluye que ambos conjuntos de datos son estadísticamente diferentes.

11. Un segundo método para realizar el mismo ensayo es calcular el intervalo de confianza ( $CI$ ) en relación con la diferencia entre los dos valores medios con arreglo a la siguiente fórmula:

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

donde  $\bar{x}_1$  y  $\bar{x}_2$  son los valores medios de cada conjunto de muestras de cada método;  $t_{95,(n_1+n_2-2)}$  es el valor  $t$  con un nivel de confianza del 95 % con  $n_1 + n_2 - 2$  grados de libertad;  $s_p$  es la estimación de la desviación estándar combinada para los conjuntos de muestras con arreglo a la ecuación 2);  $n_1$  y  $n_2$  representan el número de muestras objeto de análisis de cada conjunto. El valor de  $t_{95,(n_1+n_2-2)}$  puede encontrarse en tablas estadísticas o utilizando la función T.INV de Excel. Si  $CI$  no incluye un valor cero, entonces se concluye que los promedios son considerablemente diferentes.

12. Se ha señalado que el ensayo  $t$  para dos muestras tal vez no sea apropiado para determinar la equivalencia entre dos métodos, puesto que se ha concebido para demostrar que dos conjuntos de datos tienen promedios diferentes y no que tienen promedios equivalentes<sup>6, 8, 9, 10, 11</sup>. Puesto que la hipótesis nula o por defecto es que los promedios son iguales, el método proporciona pruebas de apoyo únicamente si los dos promedios son diferentes. Cuando no existen pruebas suficientes para demostrar una diferencia significativa en los promedios, se acepta por defecto la hipótesis nula, pero ello no demuestra que los promedios sean realmente iguales. Se han determinado dos problemas específicos relacionados con la aplicación del ensayo  $t$  de dos muestras al evaluar la equivalencia de métodos. En primer lugar, cuando las desviaciones estándar de los dos conjuntos de datos son relativamente elevadas y el número de mediciones es relativamente bajo, el ensayo  $t$  puede indicar que no hay diferencias significativas, mientras que al incluir mediciones adicionales se hubiera demostrado que los promedios eran de hecho estadísticamente diferentes. En segundo lugar, cuando las desviaciones estándar para los dos conjuntos son relativamente pequeñas, el ensayo puede tener como resultado una diferencia estadística aunque la diferencia absoluta sea prácticamente insignificante.

### Opción 2: Dos ensayos $t$ unilaterales (TOST)

13. Como alternativa recomendada al ensayo  $t$  de dos muestras se ha propuesto un contraste estadístico de dos ensayos  $t$  unilaterales (TOST) que parte de la hipótesis nula contraria, es decir, que los dos valores medios de los métodos en cuestión no son equivalentes<sup>6, 7, 11</sup>. Por tanto, un contraste de significación positivo, a un determinado nivel de confianza, demuestra que los dos conjuntos de datos son equivalentes. El ensayo TOST exige la especificación de un parámetro denominado criterio de aceptación,  $\pm\theta$ , que representa la diferencia más pequeña en los valores medios de ambos métodos considerada prácticamente importante. La hipótesis nula,  $H_0$ , y la alternativa,  $H_a$ , se describen en términos de la diferencia en los promedios,  $\mu_1 - \mu_2$ , y  $\theta$ , mediante las siguientes fórmulas:

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

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

14. La hipótesis alternativa se demuestra a un determinado nivel de confianza, cuando la verdadera diferencia en los promedios entre métodos se encuentra dentro de los límites determinados por  $\pm\theta$ .

15. El ensayo TOST se realiza especificando en primer lugar la variable  $\theta$  y calculando después el rango de confianza ( $CI$ ) respecto a la diferencia en los promedios a un determinado nivel de confianza (normalmente del 95%) mediante la siguiente fórmula:

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

donde  $\bar{x}_1$  y  $\bar{x}_2$  son los valores medios de cada conjunto de muestras de cada método;  $t_{90,(n_1+n_2-2)}$  es el valor  $t$  con un nivel de confianza del 90% con  $n_1 + n_2 - 2$  grados de libertad;  $s_p$  es la estimación de

la desviación estándar combinada para los conjuntos de muestras con arreglo a la ecuación 2);  $n_1$  y  $n_2$  corresponden al número de muestras objeto de análisis de cada conjunto. El valor de  $t_{90,(n_1+n_2-2)}$  puede encontrarse en tablas estadísticas o utilizando la función DISTR.T.INV de Excel. (Cabe señalar que el límite de confianza unilateral del 95 % es equivalente al límite de confianza bilateral del 90 %). Si  $CI$  queda completamente comprendido en el rango definido por  $\pm\theta$ , los métodos definidos por este conjunto de datos se consideran equivalentes.

16. Ambos problemas relacionados con el ensayo  $t$  de dos muestras se resuelven utilizando el enfoque TOST debido a la especificación del criterio de aceptación,  $\theta$ . La situación en que se determine una diferencia estadística en los promedios que es prácticamente insignificante queda resuelta puesto que la especificación del rango ( $\pm\theta$ ) permite un sesgo estadísticamente significativo siempre y cuando sea lo suficientemente pequeño como para quedar comprendido en el rango. Se tiene en cuenta también la situación en que los valores  $s$  del método son relativamente amplios, puesto que no se establecerá la equivalencia del método salvo que el  $CI$  calculado para los dos métodos quede comprendido íntegramente en el rango definido por  $\pm\theta$ .

17. Ello conduce asimismo a lo que podría percibirse como una desventaja del enfoque TOST: la necesidad de especificar el criterio de aceptación. Puede resultar difícil establecer un solo criterio de aceptación para todos los métodos, ya que ello requiere determinar la menor diferencia prácticamente significativa en los valores medios y puede exigir una evaluación individual de cada método.

18. Para obtener la potencia o la confianza adecuadas como conclusión de un contraste de significación estadística es necesario analizar un número suficiente de muestras. El número de muestras necesarias depende de la ratio entre  $\theta$  y  $s_p$ ; una ratio mayor requiere un número menor de muestras. Cuando los valores  $\theta$  y  $s_p$  son similares, se necesitan entre 18 y 27 muestras para obtener un nivel adecuado de potencia respecto al ensayo de equivalencia<sup>9</sup>.

### Opción 3: Límites de concordancia

19. Una tercera opción para evaluar la equivalencia consiste en adoptar un enfoque basado en el cálculo de los límites de concordancia ( $LOA$ ) en torno al promedio del método de Tipo I aprobado<sup>2,3,4,5</sup>. Para cada conjunto de datos analizado, el  $LOA$  calculado representa el rango en torno al promedio en que cabría encontrar el 95% de los resultados. Al igual que en otros procedimientos, se presupone una distribución normal y un número suficiente de muestras a fin de proporcionar estimaciones razonablemente buenas para los valores de los promedios y de la desviación estándar de cada método.

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

donde  $u$  es el promedio,  $z_{95}$  es el percentil 95 % de la distribución normal y  $\sigma$  es la desviación estándar para el conjunto de datos del método de Tipo I aprobado.

20. El  $LOA$  sería por tanto el rango de referencia al comparar los distintos valores del método alternativo. Si el 95 % de estos valores se encuentra dentro de los límites de concordancia, el método alternativo se consideraría equivalente respecto a ese conjunto de datos en particular. A fin de obtener un grado suficiente de confianza para este ensayo en concreto, se recomienda analizar un mínimo de 20 muestras por cada conjunto de datos puesto que 19/20 en este rango constituirían el 95% de los resultados. El enfoque de límites de concordancia permitiría un sesgo entre métodos, pero solo si es lo suficientemente pequeño como para que el 95 % de los resultados se encuentre en el  $LOA$  definido por el método de Tipo I aprobado.

### RESUMEN

21. La ventaja del ensayo  $t$  de dos muestras es que no requiere parámetros adicionales tales como un criterio de aceptación. Sin embargo, el contraste estadístico  $t$  no es apropiado cuando lo que se desea es demostrar una equivalencia. En particular, la utilización del ensayo  $t$  podría llevar a la conclusión de que los métodos son estadísticamente distintos cuando la diferencia absoluta es prácticamente insignificante. El ensayo también puede llevar a la conclusión errónea de que dos métodos son equivalentes cuando las muestras objeto de estudio son insuficientes para demostrar la diferencia estadística real.

22. El ensayo TOST de equivalencia estadística es probablemente el enfoque más riguroso, pero requiere la especificación de un límite de aceptación práctico que represente el sesgo mínimo aceptable entre los métodos. En la práctica, puede ser difícil establecer un conjunto de criterios de aceptación para todos los métodos del Codex, ya que cada uno tiene una aplicación específica. Una opción podría consistir en el empleo de los factores de recuperación especificados en el Manual de Procedimiento del Codex, en la sección sobre “directrices para establecer valores numéricos relativos a los criterios”, con la concentración adecuada. Por ejemplo, la recuperación especificada de 95-105 % para concentraciones de  $\geq 0,1$  sugiere un criterio de aceptación de  $\pm 5\%$ . Sin embargo, muchos métodos no tendrán un factor de recuperación aplicable. Ello requeriría que el Comité del Codex sobre los Productos recomendara el método para examinar el sesgo mínimo aceptable apropiado caso por caso.

23. Una posible opción parece ser el enfoque *LOA* que comporta calcular un rango de probabilidad del 95 % en torno al promedio del método de Tipo I existente. Esta opción permite un sesgo únicamente si el método alternativo también genera una desviación estándar menor, de modo que el 95 % de los resultados sigan dentro del rango definido por el *LOA*. La ventaja de este enfoque sería que es muy fácil de calcular y que no requiere el establecimiento de criterios de aceptación. Una desventaja sería que no permitiría sesgos en el método alternativo si la desviación estándar del método alternativo es igual a la del método aprobado. Debido a la comparación de múltiples conjuntos de datos representativos de diferentes matrices y concentraciones puede resultar difícil para cualquier método alternativo demostrar la equivalencia utilizando este procedimiento.

24. Cada uno de estos posibles procedimientos requeriría el análisis de múltiples conjuntos de muestras que incluyan la matriz y los niveles de concentración adecuados para abarcar el ámbito del método. Se sugiere que se analicen por lo menos dos concentraciones para cada matriz pertinente. Una debería estar en el límite máximo o mínimo (LM) y la otra en el límite de cuantificación. Si el método no tiene un LM especificado, debería elegirse una concentración baja y otra alta que abarquen el intervalo de concentraciones del método. Respecto al procedimiento TOST, el número de muestras que habría de analizarse para demostrar la equivalencia varía en función de la desviación estándar y el criterio de aceptación, pero una estimación conservadora de 18 a 27 muestras para cada conjunto de datos abarcaría situaciones donde  $\theta$  es similar en magnitud a  $s_p$ . Se recomiendan al menos 20 muestras para el enfoque *LOA* con el fin de obtener una estimación razonablemente buena de la desviación estándar. Los estudios de equivalencia deberían realizarse, en la medida de lo posible, en condiciones de reproducibilidad. Podrían llevarse a cabo en un solo laboratorio, aunque sería conveniente utilizar, en la medida de lo posible, analistas, materiales y equipo diferentes.

#### INTERROGANTES PARA EL DEBATE

25. Aunque en el documento se describe en general una serie de enfoques estadísticos para establecer una equivalencia entre métodos, es evidente que no se abordan detalles específicos sobre la aplicación de ninguno de estos métodos. Si bien puede recomendarse un enfoque general, todavía hay interrogantes pendientes sobre la aplicación de un tal enfoque a métodos específicos. Por ejemplo,

- i. Considerando lo expuesto y la gama de métodos existente ¿es práctico establecer un conjunto de criterios de equivalencia para todos los métodos del Codex?
- ii. Si se establecieran dichos criterios, o incluso procedimientos generales para evaluar la equivalencia ¿dónde se ubicarían en el Codex, como parte del Manual de Procedimiento o en un documento de orientación?

#### Referencias

1. Freund, J. E.; *Modern Elementary Statistics*, 7.<sup>a</sup> ed.; 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. Schuirman, 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.