

## Homogeneidad en la evaluación de procedimientos de medida

A. González Lartitegui

La evaluación de procedimientos de medida es un capítulo de la química clínica que genera un gran número de publicaciones en las revistas especializadas. Sin duda, dos factores responsables de la fecundidad de artículos en este terreno son, por un lado, la facilidad de diseño y por otro, el carácter descriptivo del estudio, que no implica dificultades *a priori* en la interpretación de los resultados. Además, las conclusiones derivadas de tales estudios, tienen una aplicación práctica inmediata. El primero de estos factores es fruto de la disponibilidad de diversos protocolos que describen detalladamente los aspectos técnicos que debe cubrir una evaluación (1-4); dichos protocolos existen gracias a esfuerzos invertidos por organizaciones internacionales, comisiones de expertos y algunas asociaciones nacionales de química clínica para apoyar la selección y evaluación de procedimientos de medida en unos criterios objetivos. Una vez obtenidos los resultados mediante un análisis estadístico adecuado, éstos deberían compararse con unos ideales de calidad, definidos también por grupos de expertos internacionales (5,6) y basados en consideraciones teóricas, que permiten eliminar la subjetividad en su interpretación.

A pesar de esta lógica tendencia a unificar criterios, se puede apreciar en la literatura «químico-clínica» que no siempre se emplean los protocolos recomendados por las asociaciones internacionales, y se siguen criterios basados en la propia experiencia, por lo que las conclusiones obtenidas son, en estos casos, discutibles. Se observa además una notable disparidad (y, en ciertos casos, ausencia) de criterios en el análisis estadístico e interpretación de los resultados. No obstante, estos protocolos pueden resultar algo teóricos y, en algún caso particular, ser difícilmente practicables. El motivo de esta carta es revisar brevemente dos de las revistas españolas del ámbito del laboratorio clínico «Química Clínica» y «Revista de Diagnóstico Biológico» y, en particular, los artículos y cartas técnicas publicados durante 1992 en que se describen evaluaciones de procedimientos de medida.

Los protocolos descritos generalmente recomiendan la valoración de la imprecisión, la inexactitud, la detectabilidad, la especificidad analítica y la practicabilidad del procedimiento de medida, así como la definición del intervalo analítico como principales objetivos de una evaluación. También se recomienda valorar la transferibilidad de resultados entre el procedimiento en estudio y el procedimiento empleado habitualmente, si existe, o bien con respecto a un procedimiento de conocida inexactitud, a fin de garantizar la transferibilidad de resultados. De los once estudios revisados (7-17) en cinco se estimó la imprecisión del método a partir de un número de observaciones inferior al recomendado y, en dos de ellos, se determinó la magnitud en estudio en un solo espécimen (diversas organizaciones interna-

cionales recomiendan un mínimo de tres (1,2)). En cuanto a la estimación de la inexactitud, sólo en dos artículos los autores comparan el valor obtenido por el procedimiento en estudio con un valor asignado. No obstante, en uno de ellos se utiliza el material de control incluido en el equipo, cuyo valor se asignó con el mismo procedimiento a valorar, por lo que carece de sentido tal comparación. De los nueve artículos en que se realiza la prueba de recuperación, en cuatro de ellos se realiza añadiendo materiales que no están debidamente valorados (como materiales de control interno o especímenes humanos valorados con el mismo procedimiento en estudio) a especímenes humanos, con lo que el valor teórico está afectado por un error sistemático desconocido. En general, se recomienda aumentar la concentración del constituyente en estudio a partir de la sustancia pura o de un material valorado por un método de referencia. En seis de los estudios se estimó el límite de detección del procedimiento, calculado en todos los casos a partir de resultados de una misma serie analítica, cuando debería calcularse a partir de determinaciones hechas en diferentes series (18); además, en tres estudios se emplea la fórmula  $\bar{x} \pm 2,33s$  para su cálculo, mientras que en los otros tres, se calcula a partir de la fórmula  $\bar{x} \pm 3s$ . En los otros cinco estudios no se determinó el límite de detección. La practicabilidad de un procedimiento de medida abarca aspectos que habitualmente la literatura técnica se encarga de describir, y otros aspectos menos evidentes pero no menos importantes como es la disponibilidad de materiales de control adecuados. En las evaluaciones revisadas, este apartado se cubre de una forma irregular, no incluyendo, por ejemplo, estudios de coste.

Existe también cierta confusión en el tratamiento estadístico de los resultados, particularmente para el estudio de la transferibilidad de resultados entre diferentes procedimientos. Es frecuente encontrarse con estudios en los que se valora la transferibilidad de resultados entre dos procedimientos, teniendo en cuenta exclusivamente el coeficiente de correlación lineal, cuando este estadístico únicamente refleja el grado de asociación entre dos variables (19). La medición de la misma magnitud por dos procedimientos de medida diferentes debe tener un alto grado de asociación, pero no implica la intercambiabilidad entre dichos procedimientos. Por ello es necesario calcular la ecuación de regresión y valorar la proximidad de sus parámetros a los de la recta de identidad,  $y=x$ . Además, las peculiaridades estadísticas de las poblaciones de datos a enfrentar hacen que no sea correcto aplicar la regresión lineal simple para este tipo de estudios (20), siendo preferible utilizar la regresión ortogonal de Deming o la regresión no paramétrica de Passing y Bablok. En seis de los estudios revisados se valoró la transferibilidad de los resultados utilizando la regresión de Passing y Bablok (si bien en uno de ellos con un número de datos insuficiente), en otros dos se empleó la regresión lineal sin calcular los intervalos de confianza de los parámetros estimados, y en tres estudios únicamente se calculó el coeficiente de correlación lineal.

A. González Lartitegui  
Progreso, 76, ático  
08904 L'Hospitalet de Llobregat  
Barcelona

Por lo que respecta a la interpretación de los resultados, la ausencia de criterios objetivos que establezcan los límites tolerables de imprecisión e inexactitud de un procedimiento es la deficiencia más llamativa. Así, en tan sólo tres de los once trabajos revisados se compara la imprecisión obtenida con la imprecisión máxima admisible calculada a partir de datos de variabilidad biológica, uno de los criterios más ampliamente aceptados para valorar la imprecisión de un procedimiento, aunque no el único. En la actualidad se dispone de abundante información acerca de la variabilidad biológica de numerosas magnitudes biológicas, lo que permite contrastar datos sin dificultad (21,22). Asimismo, prácticamente en la totalidad de artículos, la inexactitud del procedimiento evaluado no se compara con unos límites tolerables de inexactitud; sin embargo, en algunos de ellos se concluye que el procedimiento tiene una inexactitud «aceptable».

En conclusión, se percibe cierta falta de información por parte de los grupos que realizan las evaluaciones, que debería ser evitada, ya que en algunos casos los resultados del proceso de evaluación transmiten cierto grado de escepticismo que hace que se multiplique el trabajo y el dinero invertidos en la evaluación de un procedimiento de medida.

#### Bibliografía

1. International Federation of Clinical Chemistry. Approved recommendation (1978) on quality control in clinical chemistry. Part 2. Assessment of analytical methods for routine use. *J Clin Chem Clin Biochem* 1980; 18: 78-88.
2. European Committee for Clinical Laboratory Standard. Guidelines for a user laboratory to evaluate and select a kit for its own use. ECCLS document. Berlin: Beuth Verlag GmbH, 1986. Vol 3, N° 3.
3. White GH, Fraser CG. The evaluation kit for clinical chemistry: a practical guide for the evaluation of methods, instruments and reagent kits. *J Autom Chem* 1984; 6, 122-141.
4. Commission «For validation of methods» of the Société Française de Biologie Clinique. Protocol for the validation of methods (document B, stage 3). *Ann Biol Clin* 1986; 44: 686-745.
5. Fraser CG. Desirable performance standards for clinical chemistry tests. *Adv Clin Chem* 1983; 23: 299-399.
6. Gowans EMS, Peterson PH, Blaabjerg O, Morder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988; 48: 757-764.
7. Ras RM, Riera J. Evaluación de un método inmunofelométrico para la determinación de microalbuminuria. *Rev Diagn Biol* 1992; 41: 53-56.
8. Esteban B, Iglesias A, Ruiz R, Borque L. Determinación inmunoturbidimétrica de Cl-esterasa inhibidor de un analizador centrifugo. *Rev Diagn Biol* 1992; 41: 418-423.
9. Jaqueti J, Alonso E, Muñoz C, Navarro-Gallar F. Validación del método LAPIA para cuantificación de ferritina. *Rev Diagn Biol* 1992; 41: 387-388.
10. Jaqueti J, Contreras R, Cuesta G, Navarro-Gallar F. Validación del método LAPIA para cuantificar fracciones C3 y C4 del complemento. *Rev Diagn Biol* 1992; 41: 330.
11. Alcaraz A, Noguera J, Tovar I, Martínez P. Estudio de la medida de cortisol libre en orina (UFC) por inmunoanálisis de polarización fluorescente (FPIA). Correlación con la determinación de 17-hidroxiesteroides. *Rev Diagn Biol* 1992; 41: 411-413.
12. Ras RM, Soriano B. Evaluación de un método inmunoenzimático para la determinación de la concentración de ferritina sérica en el analizador Abbott IMx®. *Quím Clín* 1992; 11: 171-175.
13. Esteban Salan M, Amoroto E, Izquierdo Quirze E, Mar Medina C, López-Urrutia A. Determinación de prolactina por enzoinmunoanálisis de partición radial. *Quím Clín* 1992; 11: 422-428.
14. Espárrago Rodilla M, Baz Alonso MJ. Evaluación de las determinaciones de tiroxina y capacidad de fijación de tiroxina realizadas mediante el método CEDIA® en un analizador Hitachi 717. *Quím Clín* 1992; 11: 429-434.
15. Navarro Navarrete C, Pena Ezquerro JM, Trapé i Pujol J, Aulesa Martínez C, Botey i Sala J, Sentís Vilalta M. Determinación de la concentración de inmunoglobulina E por enzoinmunoanálisis de captura de micropartículas. *Quím Clín* 1992; 11: 439-443.
16. Martínez Compadre G, De No Lenganan C, Aniel-Quiroga Rodríguez MA, Busturia Jimeno MA. Cuantificación de la tirotrona y de la triyodotironina séricas por un método inmunoenzimático de captación de micropartículas. *Quím Clín* 1992; 11: 17-23.
17. Ortolá J, Arranz B, Rosel P, Bonnin MR, Navarro MA. Evaluación de la calidad analítica de las magnitudes hormonales de función tiroidea, determinadas mediante métodos ELISA aplicados al analizador ES-600 Enzymun-Test®. *Quím Clín* 1992; 11: 24-28.
18. Fuentes-Arderiu X. Detectabilidad, límite de detección y sensibilidad analítica. Educación continuada en química clínica 1990; 3: 1-6.
19. Domènech JM. Bioestadística. Métodos estadísticos para investigadores. Barcelona: Herber, 1980.
20. Frey E, Fuentes-Arderiu X, Queraltó JM. Comparación estadística de métodos analíticos. Educación continuada en química clínica 1988; 1, 69-80.
21. Fraser CG. Biological variation in clinical chemistry. *Arch Pathol Lab Med* 1992; 116: 916-923.
22. Fraser CG, Hyltoft Petersen P, Ricós C, Haecckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem* 1992; 30: 311-317.