

Utilización adecuada de las pruebas estadísticas en los estudios de comparación de procedimientos

F. López Azorín

Sr. Director:

Siendo la comparación estadística de los resultados de un método analítico respecto a los obtenidos por un método de referencia, una de las etapas imprescindibles en todo proceso de validación para conocer la exactitud del mismo, la importancia del problema planteado en la carta de Caballero Sarmiento (1) al dudar de la idoneidad de las técnicas de regresión lineal y esperar nuevos estudios que establezcan su validez, obligan a comentar las últimas recomendaciones publicadas por su trascendencia en la calidad de nuestros resultados.

Durante años los métodos de regresión lineal han sido los más utilizados e incluso adoptados como estándar en la revista *Clinical Chemistry* donde además de requerir un correcto análisis estadístico de la regresión, señalan la limitada utilidad del coeficiente de correlación (2). En este sentido Altman et al. en 1983 afirmaban que el coeficiente de correlación «es uno de los estadísticos peor utilizados» y que «es inapropiado para comparar procedimientos alternativos de medida de la misma variable porque valora asociación pero no concordancia» (3). Recientemente la divulgación de la prueba de comparación de medias publicada por Bland y Altman en 1986 (4) y la defensa de su mayor idoneidad frente a la correlación lineal (5) ha motivado una polémica que obliga a conocer los estudios críticos sobre las condiciones idóneas de aplicación de ambas pruebas estadísticas (6,7).

En esta situación tiene especial importancia el editorial publicado en noviembre de 1998 por Westgard en *Clinical Chemistry* estableciendo diez puntualizaciones para su adecuada utilización e intentar zanjar la polémica (8). Partiendo de las conclusiones de su estudio de 1973 al respecto (9), donde ya apuntaba que los estadísticos r , t o F no permiten establecer la aceptabilidad de un procedimiento, sino que deben proporcionar la estimación del error, recuerda que debe conocerse si el error sistemático del procedimiento es menor que el máximo error permisible sin comprometer significativamente la utilización clínica de los resultados. Ello exige que las concentraciones de las muestras del estudio deben cubrir las concentraciones de decisión clínica del constituyente correspondiente.

El esquema de decisiones propuesto por Westgard parte de conocer el número de concentraciones de decisión clínica del constituyente. Si hay una sola concentración, la estimación del error sistemático puede realizarse tanto con regresión lineal como con la prueba de t . Si hay dos o más concentraciones de decisión, la estimación que proporcionan ambos estadísticos es diferente y recomienda utilizar el coeficiente de correlación r para decidir el tipo de prueba. Así mediante r se juzga si el intervalo de concentraciones utilizado permite aplicar el cálculo de la regresión. Si r es alto ($r > 0,990$) la regresión lineal permite la estimación del error, mientras que si es bajo ($r < 0,975$) la regresión lineal simple no es aceptable para ese conjunto de datos y debe mejorarse la calidad de los datos o su intervalo de concentraciones, o bien aplicar técnicas de regresión más rigurosas -v. gr. Passing-Bablok o me-

por la de Deming- o también la prueba de comparación de medias aplicando como estadístico la prueba t , que mide la proporción entre error sistemático y aleatorio y valora si los datos recogidos son suficientes para estimar el sesgo, pero, análogamente a r , tampoco permite estimar la aceptabilidad del procedimiento. Para aplicar el método de comparación de medias es imprescindible representar gráficamente las diferencias y comprobar que la diferencia de resultados entre ambos métodos muestra un error sistemático constante, pues tal como señaló Linnet esta prueba no puede ser aplicada si el error es proporcional (10).

Sin embargo estas recomendaciones no garantizan la solución del problema en todos los casos, pues Westgard -además de no indicar la decisión cuando los valores de r quedan excluidos de los dos grupos- en sus últimas puntualizaciones aconseja que puede ser necesario comparar los resultados con diferentes pruebas estadísticas, extremando la cautela cuando la discrepancia compromete la decisión de aceptar un método, como es el caso planteado en la carta de Caballero. Ello obligaría a replantear el experimento de comparación, extremando la rigurosidad en la selección de muestras y obtención y análisis de resultados según las exigencias de la prueba estadística a utilizar conforme a las recomendaciones señaladas (6, 7, 10).

A las dificultades reseñadas debe añadirse la necesidad de conocer los límites de decisión clínica del constituyente cuyo procedimiento de medida se compara, pues aunque están recopilados en la última edición de 1987 de la obra de Statland (11), su falta de actualización es patente en magnitudes como colesterol de HDL, apolipoproteínas, isoenzimas o ferritina. Considerando la importancia de las concentraciones de decisión tanto en la comparación de procedimientos, como en la adecuada selección de los materiales de control de la calidad, sería conveniente que las correspondientes comisiones de las sociedades científicas los proporcionasen en sus recomendaciones de la forma más actualizada posible.

Bibliografía

1. Caballero Sarmiento R. ¿Son adecuadas las técnicas de regresión para comparar procedimientos analíticos? *Quim Clin* 1999; 18: 288.
2. Information for contributors. *Clin Chem* 1996; 42: 1-4.
3. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J* 1983; 286: 1489-93.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-10.
5. Hollis S. Analysis of methods comparison studies [editorial]. *Ann Clin Biochem* 1996; 33: 1-4.
6. Hyltoft Petersen P, Stockl D, Blaabjerg O, Pedersen B, Birkemose E, Thienpont L et al. Graphical interpretation of analytical data from comparison of a field method with a reference method by use of difference plots [opinion]. *Clin Chem* 1997; 43: 2039-46.
7. Stockl D, Dewitte K, Thienpont M. Validity of linear regression in method comparison studies: limited by the statistical model or the quality of the analytical data? *Clin Chem* 1998; 44: 2340-6.
8. Westgard JO. Points of care in using statistics in method comparison studies. *Clin Chem* 1998; 44: 2240-2.
9. Westgard JO, Hunt MR. Use and interpretations of common statistical test in method comparison studies. *Clin Chem* 1973; 19: 49-57.
10. Linnet K. Limitations of the paired t-test for evaluation of method comparison data [letter]. *Clin Chem* 1999; 45: 314-5.
11. Statland BE. Clinical decision levels for lab test. Second ed. Oradell NJ: Medical Economics Books, 1987.