Estándar ideal imperfecto y variable

Gastón Duffau T.¹

Faulty and variable comparison standards

The operative aspects of a diagnostic test are usually defined by a gold or ideal comparison standard, which is expected to accurately separate subjects with and without the investigated anomaly and also to set up important indicators of the test like its sensitivity, specificity and predictive values. However, these indicators may show great variability from one trial to another one depending on the prevalence, spectrum and presence in the sample of the abnormality under observation. Ideal standards are seldom perfect even though they are frequently treated as if they were. Thus they may generate variable classifications of cases, which is also dependent on the spectrum of the observed phenomena in the sample. In front of such serious difficulties in the selection of appropriate spectrum samples, interpretation of results should be done with extreme caution, particularly when newly developed diagnostic tests are evaluated.

(Key words: epidemiological method, statistics.)

En el estudio de una prueba diagnóstica, su comparación con el estándar escogido —que nos permitirá separar los que tienen la alteración buscada de aquellos que no la padecen— caracterizará a aquélla con dos índices considerados invariables (sensibilidad y especificidad), así como con dos índices esencialmente cambiantes, según la prevalencia (de acuerdo con el estándar) del trastorno investigado (valores predictivos, positivo y negativo).

Parece claro que los índices “fijos” pueden, aunque no necesariamente lo hacen, mostrar inestabilidad al variar la prevalencia (Sackett)¹ aunque a veces lo hacen en forma muy notoria ante una misma prueba diagnóstica efectuada por grupos distintos de investigadores². Por otra parte, el espectro de la alteración investigada, debe estar adecuadamente representado en la muestra analizada, porque tanto la sensibilidad como la especificidad de la prueba pueden ser diferentes en cada subgrupo³. Puesto que, en el conjunto que incluye todo el espectro, sensibilidad y especificidad son en definitiva una suerte

de valor “promedio” de los obtenidos en los subgrupos integrantes, es muy importante la representación numérica que tenga cada subgrupo en el conjunto, no bastando simplemente con que exista “un espectro” que contenga todas las variantes cualitativas. La necesidad de dar una solución al problema obliga a aceptar algunos supuestos, como que para cada prueba diagnóstica subyace una sensibilidad y una especificidad fijas, a las que es posible acercarse asintóticamente en la medida que se cumplan los requisitos que exige el ensayar una prueba diagnóstica¹. El espectro adecuado cualitativamente y cuantitativamente debería ser, en definitiva, una muestra representativa, de tamaño adecuado, del respectivo universo estadístico. Para el caso, se avizora como una tarea muy difícil de efectuar bien, aun sin considerar que tal universo debería incluir la situación patológica en su forma subclínica.

Por otro lado, al estándar de comparación o patrón oro, una vez elegido, se le trata como perfecto, aunque no lo sea. Es posible encontrar interesantes consideraciones relativas a los efectos esperables en los resultados de una prueba diagnóstica cuando el patrón es imperfecto⁴. Sin embargo, no se contempla que éste pueda, además, generar una clasificación variable de la

1. Departamento de Pediatría, Cirugía Infantil y Ortopedia Norte, Facultad de Medicina, Universidad de Chile. Hospital de Niños Roberto del Río.

muestra en análisis (proporción de pacientes con y sin el trastorno de interés). Dicha variabilidad puede depender también del espectro en la muestra del trastorno que se investiga. Así, por ejemplo, al tomar el caso de la infección urinaria, veremos que si el patrón de comparación es el recuento de colonias en el urocultivo, clasificando los casos como positivos si muestran más de 100 000 UFC/ml y negativos los restantes, se podrá reconocer que la probabilidad de encontrar verdadera infección del tracto urinario en el grupo con menos de 100 000 UFC/ml, variará con edad, sexo, antecedentes previos de la afección, manifestaciones clínicas, etc. Si nuestro problema es calificar estado de hidratación, empleando como estándar la "ganancia de peso" una vez hidratado el paciente, deberemos reconocer que el patrón de comparación operará francamente peor en los niños con deshidratación por diarrea, en que exista un tercer espacio abdominal importante, situación donde es posible —y aun probable— que el paciente se hidrate sin cambio notorio del peso corporal. Mejor resultado sería de esperar en aquéllos sin franco bazuqueo al examen físico así como ante deshidratados por causas no digestivas.

No resolveremos aquí este problema pero señalaremos entonces que el espectro puede afectar no sólo las condiciones de operación de una prueba sino también del estándar cuando éste es imperfecto. Esto constituye una razón adicional para ponderar aún más cuidadosamente los resultados de las pruebas diagnósticas.

Resumen

Las características de operación de una prueba diagnóstica son definidas mediante un patrón estándar ideal de comparación, del que se espera sea capaz de separar acertadamente a quienes padecen y no la anomalía investigada y, además contribuir a establecer importantes índices de la prueba, como sensibilidad, especificidad y valores predictivos. Estos índices resultan ser —en último término— variables, entre uno y otro ensayo, de acuerdo con la prevalencia, el espectro y la representación de la alteración investigada en la muestra. El patrón ideal es pocas veces perfecto, aunque se le trate como tal, y puede generar clasificación variable de los casos, lo que también depende del espectro del trastorno en la muestra. Ante las serias dificultades para estructurar conjuntos de espectro apropiado, es preciso tener extrema precaución al interpretar los resultados obtenidos, especialmente cuando se aplican nuevas pruebas de diagnóstico.

(Palabras clave: método epidemiológico, estadísticas.)

Referencias

1. Sackett, Haynes B, Guyatt, Tugwell P: Clinical Epidemiology, 2nd Edition, Little, Brown and Co., Boston 1991: 88.
2. Harris J: The hazards of bedside Bayes. JAMA 1981; 246: 2602-2605.
3. Herrera P, Lagos R: Contribución al estudio de pruebas diagnósticas: sobre sus índices fijos y valores predictivos. Rev Chil Pediatr 1995; 66: 209-216.
4. Valenslein P: Evaluating diagnostic tests with imperfect standards. AJCP 1990; 93: 252-258.

AVISO A LOS AUTORES

Por acuerdo del Comité Editorial, la Revista Chilena de Pediatría devolverá sin tramitar todos los trabajos que no den estricto cumplimiento al Reglamento de Publicaciones y a las Instrucciones de los Autores que se editan en cada número de la Revista.