

## Algunas consideraciones sobre la interpretación de resultados de estudios de pruebas diagnósticas: estudio DINO

### Sr. Editor:

Hemos leído con interés el trabajo de Huerta et al<sup>1</sup>. Consideramos que validar la precisión de cuestionarios de diagnóstico referido es importante desde el punto de vista epidemiológico por sus ventajas (comodidad, tiempo y coste reducido) frente a métodos biométricos y clínicos, pero nos gustaría hacer algunas consideraciones sobre dicho estudio con el fin de contribuir al debate científico y al intercambio de ideas:

1. Aunque los autores afirman que «los entrevistadores estaban bien cualificados», creemos que la utilización de algún procedimiento (p. ej., el índice kappa) para establecer la fuerza de concordancia entre ellos habría contribuido a una mayor fiabilidad del estudio<sup>2</sup>.

2. En cuanto a la prevalencia de las distintas enfermedades, los autores afirman que «todas las prevalencias referidas son inferiores a las estimadas por el patrón de referencia». Pero, en el caso de la diabetes, los intervalos de confianza (IC) del 95% del patrón y del cuestionario referido se solapan, por lo que no se puede afirmar que las medidas puntuales halladas sean diferentes entre sí.

3. En el artículo se afirma, basándose en el cálculo del índice kappa, que «se ha obtenido un grado de acuerdo entre el diagnóstico referido y el patrón de referencia bueno en diabetes, moderado en HTA y bajo en hiperlipemia». Esto no es correcto ya que, aunque el uso menos conocido del índice kappa es para valorar discrepancias de pruebas diagnósticas, el requisito indispensable para esta utilidad del índice es que la interpretación de las pruebas esté sujeta al factor humano (p. ej., radiografías, electrocardiogramas, muestras de anatomía patológica, etc.)<sup>3</sup>, lo que no ocurre en este caso, ya que las medidas del patrón de referencia son biométricas, calculadas por aparatos automáticos, calibrados y no sometidos a variabilidad humana en cada medida. Además, se compara el diagnóstico declarado (que es la suma de los datos aportados por varios investigadores no sometidos previamente a ninguna prueba de concordancia entre observadores) con los de la prueba de referencia (que es objetiva desde el punto de vista de variabilidad en cada determinación), por lo que no puede realizarse tal comparación. Además, como las sensibilidades de las tres enfermedades son inferiores al 70%, se sabe, por modelos matemáticos, que con sensibilidades infe-

riores a este 70% el índice kappa va a resultar necesariamente  $< 0,4$ , porque la baja sensibilidad conduce por sí misma a discordancias<sup>2</sup>.

4. Dado que el estudio trata de valorar la utilidad de una prueba diagnóstica (cuestionario de diagnóstico referido) tomando como estándar los diagnósticos a partir de pruebas biométricas, consideramos que, además del cálculo de sensibilidad, especificidad y valores predictivos, se debería haber calculado los coeficientes de probabilidad positivo (CPP) y negativo (CPN), ya que los valores de sensibilidad y especificidad, a pesar de definir completamente la validez (grado en que un test mide lo que se supone que debe medir) de la prueba diagnóstica, presentan la desventaja de que no proporcionan información relevante a la hora de tomar una decisión clínica ante un determinado resultado de la prueba. Los CPP y CPN expresan un resumen unificado de la sensibilidad y la especificidad y, por lo tanto, no dependen de la prevalencia de la enfermedad en cada lugar y permiten la comparación entre estudios diferentes, a diferencia de los valores predictivos, que sólo valen donde han sido calculados<sup>4,5</sup>. En este estudio, los CPP y CPN con sus IC del 95% para cada enfermedad (CPP para diabetes, 161,46 [72,21-361,06]; HTA, 15,41 [10,84-21,9]; hiperlipemia, 10,98 [7,07-17,05]; CPN para diabetes, 0,30 [0,24-0,38]; HTA, 0,52 [0,48-0,57]; hiperlipemia, 0,68 [0,64-0,71]) indican que el cuestionario es útil para diagnosticar diabetes e hipertensión (CPP  $> 10$  con IC del 95% relevantes), pero no son útiles para descartarlas (los CPN no son  $< 0,1$ ). En cuanto a la hiperlipemia, sería necesario plantear más estudios (coeficientes de probabilidad estadísticamente significativos pero no concluyentes clínicamente).

5. Cuando nos referimos a estudios de pruebas diagnósticas, se consideran parámetros óptimos o elevados si son  $\geq 80\%$ , por lo que no se puede afirmar que las sensibilidades del estudio lo sean.

En conclusión, debemos ser cuidadosos y metódicos con todos los detalles involucrados en un estudio de investigación, con el fin de contribuir de la forma más eficiente a su comprensibilidad y su utilidad para los investigadores y los clínicos a los que va dirigido.

Álvaro Campillo-Soto, Ramón Lirón-Ruiz,  
Juan Gervasio Martín-Lorenzo  
y José Luis Aguayo-Albasini

Servicio de Cirugía General y Digestiva. Hospital General Universitario J.M. Morales Meseguer. Murcia. España.

### BIBLIOGRAFÍA

- Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortolá-Devesa JB, Navarro C. Validez del diagnóstico referido de diabetes, hipertensión e hiperlipemia en población adulta española. Resultados del estudio DINO. *Rev Esp Cardiol*. 2009;62:143-52.

2. Saturno PJ. La construcción de criterios para evaluar la calidad. En: Saturno PJ, editor. Evaluación y mejora de la calidad en Servicios de Salud. Conceptos y métodos. Murcia: Consejería de Sanidad y Consumo de la Región de Murcia; 2000. p. 111-32.
3. Sackett DL. El examen clínico. En: Sackett DL, Haynes RB, Guyatt GH, Tugwell A, editores. Epidemiología clínica. Ciencia básica para la medicina clínica. Boston: Panamericana; 1997. p. 34-61.
4. Escrig-Sos J, Martínez-Ramos D, Miralles-Tena JM. Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso. *Cir Esp*. 2006;79:267-73.
5. Díaz Guzmán J. Investigación clínica. Diagnóstico. En: Diploma Superior en Metodología de la Investigación. 3.ª ed. Madrid: Instituto de Salud Carlos III; 2008. p. 1-55.

## Respuesta

### Sr. Editor:

Hemos leído y discutido con interés las consideraciones expresadas por Campillo et al en relación con nuestro artículo<sup>1</sup>. Queremos agradecer las precisiones de los autores y, especialmente, las referidas al uso del índice kappa, que ayudan a comprender y contextualizar nuestros objetivos.

Los autores hacen mención del solapamiento de los intervalos de confianza de la prevalencia de diabetes referida y la validada. En este punto, discrepamos de sus conclusiones puesto que: *a*) el solapamiento es tan sólo marginal; *b*) cada intervalo de confianza excluye la medida puntual alternativa, y *c*) tal diferencia es estadísticamente significativa (prueba de McNemar,  $p < 10^{-6}$ ).

Consideración aparte merece la precisión sobre la propiedad del uso del índice kappa como medida de acuerdo entre la información recogida mediante el cuestionario y los patrones biométricos correspondientes. Efectivamente, un estudio de estas características no es el marco habitual para la aplicación del índice, pero el valor  $\kappa$  no es el resultado más importante del análisis realizado, aunque puede que en el texto se haya destacado en exceso. Sin em-

bargo, el punto de discusión más importante es el diferente enfoque utilizado en una investigación cardiovascular poblacional como la nuestra (basada en entrevistas y en un número reducido y altamente seleccionado de pruebas biométricas realizadas una única vez para definir un diagnóstico, que puede ser individualmente inexacto pero colectivamente válido) y el centrado en la utilidad clínica en un entorno hospitalario con facilidad para solicitar y repetir un número elevado de pruebas. Nuestro objetivo era evaluar la fiabilidad de estos diagnósticos proporcionados por los encuestados en el contexto de los estudios sobre factores de riesgo cardiovascular, típicamente incluidos en estudios poblacionales transversales y de seguimiento, en los que, además, con frecuencia no son el diagnóstico de interés, sino variables de confusión necesarias para el ajuste. Este enfoque es habitual en las páginas de la propia REVISTA ESPAÑOLA DE CARDIOLOGÍA, que mantiene estable la sección «Epidemiología y prevención» entre sus páginas.

De acuerdo con nuestra orientación, la validez absoluta del cuestionario queda definida por los valores de sensibilidad y especificidad, mientras que otros índices, mostrados o no, pueden aportar información complementaria de interés variable para lectores epidemiológicos o clínicos. Consideramos que es un acierto de nuestro estudio haber aportado los datos suficientes para permitir el cálculo de índices alternativos.

Agradecemos la precisión de Campillo et al, por cuanto nos ha permitido aclarar una distinción que podría pasar inadvertida.

José M. Huerta, María-José Tormo y Carmen Navarro  
Servicio de Epidemiología. Consejería de Sanidad y Consumo  
de la Región de Murcia. Murcia. CIBER Epidemiología y Salud Pública  
(CIBERESP). España.

### BIBLIOGRAFÍA

1. Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortola-Devesa JB, Carmen Navarro. Validez del diagnóstico referido de diabetes, hipertensión e hiperlipidemia en población adulta española. Resultados del estudio DINO. *Rev Esp Cardiol*. 2009;62:15-22.