

La compleja relación entre el diagnóstico y la práctica clínica

J. I. Pijoán^a y V. Ruiz García^b

^aUnidad de Epidemiología Clínica. Hospital de Cruces. Barakaldo (Bizkaia). Miembro de la red R_MBE. Miembro de CASPe

^bUnidad de Hospitalización a Domicilio. Hospital Universitari La Fe. Valencia. Miembro de CASPe

Muchos de los recientes avances de la tecnociencia que inundan nuestros hogares y ciudades son también aplicados en nuestros centros de salud y hospitales: tarjetas sanitarias electrónicas individualizadas, archivos multimedia digitalizados, sistemas de administración parenteral continua de medicamentos en domicilio son algunas aplicaciones concretas, imposibles sin el citado desarrollo tecnológico¹. El campo del diagnóstico clínico es posiblemente uno de los que más se ha aprovechado de los avances y la gran utilidad de la información que generan algunas técnicas (mamografía, ecografía, escáner, etc.) es incluso del dominio público. Como consecuencia, el arsenal de pruebas diagnósticas a añadir a las insustituibles historia y exploración física no deja de crecer de forma continua y progresiva, paralelamente al número de publicaciones y comunicaciones que de una u otra forma valoran la utilidad de dichas pruebas².

Un aspecto que aparentemente no ha suscitado tanto interés como tema de estudio es el de la interacción entre las nuevas herramientas diagnósticas y los clínicos responsables de su uso e interpretación. Parece por tanto procedente reflexionar en voz alta sobre algunas cuestiones: ¿qué sabemos y qué debemos saber sobre las virtudes (y limitaciones) de una prueba cuando la misma aparece en el catálogo de prestaciones o existe la posibilidad (a veces la demanda) de acceder a ella? Una segunda cuestión relacionada pero con entidad propia sería: ¿cómo utilizamos este conjunto creciente de pruebas en nuestra práctica diaria y cómo las integramos de forma lógica con las herramientas disponibles?

Si observamos por un momento lo que ocurre en el campo de la terapia con medicamentos, vemos que existe un camino de "validación normativa" de los mismos necesario para su registro y comercialización. Existe un complejo entramado legal, ético y administrativo que demanda un determinado nivel de información sobre la eficacia, tolerancia y seguridad de los fármacos. Es incluso

taxativo respecto al tipo de investigación que hay que hacer: ensayos clínicos aleatorizados controlados para evaluar la eficacia relativa y estudios observacionales (de mayor duración y a veces posteriores a la comercialización del fármaco) para acumular información sobre tolerancia y seguridad. A pesar de este exigente marco de investigación terapéutica y de que existe un consenso generalizado sobre la fortaleza metodológica del ensayo clínico, abundan las críticas a la aplicabilidad real de sus resultados en la práctica cotidiana, al incluir pacientes y escenarios diferentes a los habituales. Esto es especialmente grave en Atención Primaria (donde se realizan muy pocos ensayos), al excluirse pacientes frecuentes como gestantes, ancianos, personas con alta comorbilidad o que toman ya fármacos para otros procesos³.

¿Qué ocurre en el caso de una nueva prueba diagnóstica? En función del principio que lo fundamenta (físico, químico, biológico, etc.) deberá atender determinadas exigencias técnicas de seguridad de fabricación y uso (normas ISO y de la UE). Si las cumple, quedan autorizados su registro y entrada en el mercado sanitario. A diferencia de la situación con los medicamentos, no existen unos criterios normativos exigibles de eficacia (diagnóstica) relativa o de tolerancia y seguridad. Además, tampoco está claramente determinado cuáles deben de ser los patrones de diseño de estudio para responder de forma convincente a esas preguntas. Ello probablemente contribuye a la mayor indefinición y variabilidad del campo de la investigación diagnóstica.

¿Cómo pueden los clínicos formarse una opinión clara sobre el valor de una prueba diagnóstica? Se requieren dos elementos fundamentales: por un lado, información clara proveniente de una investigación rigurosa (generalmente originales de revistas con proceso de revisión por pares) y, por otro, capacidad de interpretación crítica de esta información por parte de los profesionales, incluyendo un conocimiento operativo sobre aspectos claves del diseño de la investigación diagnóstica y de la presentación de sus resultados^{4,5}. La nota publicada por Víctor Abaira en este número⁶, junto a la publicada anteriormente⁷, son aportaciones clarificadoras en el empeño de aumentar nuestra capacidad de evaluar la investigación diagnóstica publicada.

Correspondencia: J. I. Pijoán.
Unidad de Epidemiología Clínica.
Hospital de Cruces.
Plaza Cruces s/n.
48900 Barakaldo. Vizcaya.

La forma más coherente de valorar la calidad discriminativa diagnóstica de una prueba es a través de su comparación simultánea con el mejor método (factible y éticamente aceptable) que diferencie entre la enfermedad o problema de interés y los otros estados (el llamado estándar diagnóstico de referencia o *gold standard*). El diseño lógico es de tipo observacional y transversal en la realización y comparación de los dos métodos, aunque existen varios mecanismos posibles de obtención de la muestra de estudio⁸.

Los indicadores clásicos de dicha calidad son la sensibilidad o reactividad (positividad) de la prueba ante la presencia del problema y la especificidad o ausencia de reactividad (negatividad) de la misma cuando no está presente el problema. Estos indicadores son ampliamente conocidos por los clínicos y se escuchan con cierta frecuencia en los pases de sala y sesiones médicas. La pregunta que se pretende responder con este diseño es: ¿cómo de buena es esta prueba para detectar enfermedad cuando existe y para no detectarla cuando no está presente? Por ello son estudios de la validez diagnóstica de la prueba. Como refleja el trabajo de Abaira⁶, se conoce por algunas (no muchas) investigaciones que existen aspectos del diseño de estos estudios con influencia importante en su validez y por lo tanto en la mayor o menor credibilidad que debemos dar a los mismos.

La pregunta dirigida a la validez o exactitud diagnóstica no es la única posible. Frente a este enfoque, que inevitablemente declara como error de nuestra nueva prueba cualquier resultado discordante con el estándar de referencia, existe otra posibilidad que responde a una pregunta diferente: dos métodos diagnósticos diferentes ¿nos llevan a resultados y conclusiones similares en cuanto a la presencia o ausencia de la enfermedad? El diseño es en esencia similar al anterior, pero los indicadores son diferentes: proporción simple de acuerdo entre métodos, proporción de acuerdo una vez eliminado el esperable por azar (índice kappa), etc⁹.

Todavía podemos encontrarnos estudios diagnósticos dirigidos a evaluar otras preguntas también de interés: ¿en cuántos individuos cambia la consideración de enfermo a no enfermo (o viceversa) si utilizo una nueva prueba en vez de una existente?, ¿en cuántos el cambio de diagnóstico lleva a cambios de actitud (seguimiento, tratamiento, etc.)?, ¿existe un cambio en los resultados clínicos como consecuencia de utilizar la nueva prueba en la estrategia de manejo clínico del paciente?, ¿cuánto cuesta cada diagnóstico adicional?, ¿cómo se toleran las diferentes pruebas? etc.

Por tanto, lo primero que debe estar claro es cuál era la pregunta de interés para los investigadores del estudio. Hay dos razones de peso para ello: saber si el estudio realizado es correcto para intentar responder la pregunta y conocer si la pregunta de los investigadores es la que nos interesa.

Hay un aspecto muy importante al valorar un estudio sobre la exactitud diagnóstica de una prueba: igual que en un ensayo clínico la eficacia aparente de un trata-

miento depende en gran medida de cuál sea el entorno de uso (perfil clínico del paciente y tratamiento de comparación), la exactitud aparente de la prueba variará según cuál sea el perfil clínico de los individuos del estudio y el estándar diagnóstico utilizado. Estudios que comparan los resultados de la prueba en enfermos típicos o severos frente a individuos sin enfermedad podrán ofrecer resultados (sensibilidades y/o especificidades) impresionantes pero no nos deben de interesar de forma directa si no se confirma su capacidad en entornos más reales (pacientes típicos y atípicos, pacientes con múltiples enfermedades y problemas concomitantes, etc.)¹⁰. Existen materiales de trabajo y didácticos accesibles para ayudar en la evaluación de este y otros aspectos del diseño, realización y presentación de resultados de estudios de evaluación de pruebas diagnósticas. Entre ellos destacaremos los suministrados por el Centro de Medicina Basada en la Evidencia de Oxford¹¹ y las plantillas en castellano elaboradas por el grupo español de desarrollo de habilidades de lectura crítica (CASPe)¹². Este material consigue sus mayores logros docentes cuando su uso se imbrica con la propia actividad asistencial (equipos de Atención Primaria, servicios hospitalarios, etc.) en un entorno de formación continua ceñida a problemas reales de los profesionales sanitarios.

Para que tanto la pregunta como los aspectos del diseño y procedimiento del estudio sean evaluables es obviamente necesario que sean claramente identificables en el estudio y su descripción sea lo menos ambigua posible. Desgraciadamente la situación actual está muy alejada de la deseable incluso en publicaciones que solemos utilizar como modelo¹³. Reid et al revisaron 1.300 artículos publicados a lo largo de diez años en 4 de las revistas de medicina interna general más leídas y encontraron criterios definidos de la población de estudio en apenas un 25%, esfuerzos para prevenir sesgos en sólo la mitad y reproducibilidad de las pruebas en sólo un 25%.

Recientemente un panel de expertos del grupo de trabajo Cochrane sobre métodos de cribaje y diagnóstico ha propuesto unos estándares de publicación de este tipo de estudios¹⁴ con el afán de conseguir que toda la información relevante se incluya en el artículo y sea fácilmente localizable. Hasta la fecha creemos que ninguna revista científica en castellano se ha adherido a esta iniciativa internacional, continuación de otras similares centradas inicialmente en las publicaciones sobre tratamiento.

La segunda pregunta que planteábamos se centra en la adecuada utilización de las pruebas diagnósticas en la práctica. Un buen conocimiento de las capacidades diagnósticas de las pruebas es obviamente de gran importancia para su selección, pero no garantiza que se utilicen apropiadamente. Es sabido que profesionales con mayor experiencia y grado de conocimiento clínico suelen realizar una mejor aproximación al diagnóstico a través de un diagnóstico diferencial más completo y ponderado, una selección más estructurada y eficiente de las pruebas diagnósticas y una mayor integración de la información proveniente de las mismas con el cuidado de los pacien-

tes; sin embargo los mecanismos cognitivos íntimos que rigen la gestión operativa de esta información son en gran parte desconocidos^{15,16}.

El modelo teórico de razonamiento basado en el teorema de Bayes proporciona un marco lógico muy atractivo¹⁷: el médico elabora una hipótesis diagnóstica sobre el grado de certidumbre que asigna a una enfermedad como causa del problema del paciente, basado en su conocimiento del entorno clínico concreto y tras una historia y exploración clínicas (probabilidad *a priori* o preprueba). En este momento podría ser suficiente el grado de certeza como para dar lugar a una decisión clínica concreta. Si no es así, solicitaría una prueba cuyos resultados puedan aumentar su grado de certidumbre sobre la alta o baja plausibilidad de la hipótesis diagnóstica (probabilidad *a posteriori* o postprueba) hasta llegar al punto de decisión. La información proporcionada por la prueba viene dada por los llamados cocientes de probabilidades (*likelihood ratios*) positivos y negativos según el resultado de la misma. Estos indicadores son fácilmente obtenidos a partir de la sensibilidad y especificidad⁷ y tienen la ventaja adicional de que no se ven tan afectados por la prevalencia de las enfermedades, que habitualmente es distinta en Atención Primaria de la de los centros de especialidad donde habitualmente se generan y publican.

Este modelo permite integrar la información proveniente de los resultados de las pruebas diagnósticas con su consecuencia lógica que es la acción clínica en un marco denominado análisis de la decisión clínica¹⁸. A pesar de su atractivo y de haber demostrado su utilidad¹⁵, no parece que los clínicos avezados en el diagnóstico utilicen de forma explícita este modelo de razonamiento ni sus índices asociados¹⁹, reforzando la idea de que en muchos aspectos las aproximaciones teóricas al diagnóstico clínico y a su imbricación en la acción clínica no consiguen explicar de forma satisfactoria su complejidad.

A nivel individual, sin embargo, existen dos acciones que están a nuestro alcance y por sí mismas pueden con seguridad producir efectos beneficiosos sobre el uso de las pruebas diagnósticas: exijamos por un lado claridad a los investigadores sobre la pregunta diagnóstica de interés, sobre el perfil clínico de los pacientes que incluyen en el estudio y sobre todos y cada uno de los procedimientos del mismo. La iniciativa STARD es posiblemente un primer paso en este sentido. Por otro lado, potenciemos el componente científico de nuestra profesión desarrollando la capacidad personal de valorar y criticar razonadamente la calidad y utilidad particular de la in-

formación que se nos presenta sobre diagnóstico clínico. Sólo en ese entorno podremos maximizar las posibilidades de que los verdaderos beneficiados de los grandes avances diagnósticos disponibles sean los que deben ser: nuestros pacientes.

BIBLIOGRAFÍA

1. Mesa redonda. Nuevas tecnologías y su impacto real en Atención Primaria. En: XXVII Congreso Nacional de SEMERGEN. SEMERGEN. 2005;31 Suppl 1:29-35.
2. Knottnerus JA, Van Weel C. General introduction: the evaluation of diagnostic procedures. En: Knottnerus JA, editor. The evidence base of clinical diagnosis. London: BMJ books; 2002.p.1-17.
3. Gervas J. Experiencia, ciencia y "medicina basada en pruebas" en Atención Primaria. SEMERGEN. 2002;28(6):302-4.
4. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA. 1994; 271(5):389-91.
5. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA. 1994; 271:703-7.
6. Abaira V. Sesgos en los estudios sobre pruebas diagnósticas. SEMERGEN. 2006;32(1):24-6.
7. Abaira V. Índices de rendimiento de las pruebas diagnósticas. SEMERGEN. 2002;28(4):193-4.
8. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol. 2003;56:1118-28.
9. Carrasco JL, Jover L. Métodos estadísticos para evaluar la concordancia. Med Clin (Barc). 2004;122 Suppl 1:28-34.
10. Mulherin SA, Miller C. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med. 2002;137:598-602.
11. Disponible en: <http://www.cebm.net> accedido el 3/11/2005.
12. Disponible en: <http://www.redcaspe.org> accedido el 3/11/2005.
13. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA. 1995;274:645-51.
14. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwing LM, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ. 2003;326:41-4.
15. Goldman L. Aspectos cuantitativos del juicio clínico. En: Fauci AS, Braunwald E, Isselbacher KJ, et al, editores. Harrison. Principios de Medicina Interna. 14ª ed. Madrid: McGraw-Hill-Interamericana de España; 1998.p.10-6.
16. Bloch RF, Hofer D, Feller S, Hodel M. The role of strategy and redundancy in diagnostic reasoning. BMC Medical Education. 2003; 3:1. Disponible en: <http://www.biomedcentral.com/1472-6920/3/1> accedido el 3/10/2005.
17. Abaira V. Inferencia estadística bayesiana. SEMERGEN. 2005;31(1):18-20.
18. Gross R. Making medical decisions. An approach to clinical decision making for practicing physicians. Philadelphia: American College of Physicians; 1999.
19. Reid CM, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. Am J Med. 1998;104:374-80.