OFTALMOLOGÍA BASADA EN PRUEBAS: EVALUACIÓN CRÍTICA DE LA LITERATURA SOBRE PRUEBAS DIAGNÓSTICAS

EVIDENCE-BASED OPHTHALMOLOGY: CRITICAL EVALUATION OF THE LITERATURE IN RELATION TO DIAGNOSTIC TESTS

MESA JC[1], GARCÍA O[2], LILLO J[2], MASCARÓ F[2], ARRUGA J[1]

RESUMEN

Objetivo: En general, los artículos sobre diagnóstico suelen adolecer de una calidad metodológica muy pobre. Si trasladamos sus conclusiones a nuestra práctica cotidiana sin un análisis detenido entenderemos fácilmente un fenómeno creciente: incorporamos acríticamente a nuestras estrategias diagnósticas las novísimas (y carísimas) pruebas sin que con ello aumente significativamente el rendimiento diagnóstico de tales estrategias; sólo se incrementa desorbitadamente el gasto.

La Medicina Basada en Evidencias es el uso de la mejor evidencia disponible en la toma de decisiones. Se trata de actuar utilizando información seleccionada y avalada por datos obtenidos a través del método científico más riguroso: la epidemiología y la estadística.

La evaluación crítica de artículos relacionados con pruebas diagnósticas constituirá nuestro objetivo. Suministraremos las habilidades básicas para la evaluación y análisis de artículos científicos mediante el uso de una serie de conocimientos sencillos de estadística, diseño de investigaciones y epidemiología clínica.

Método: Revisión de la literatura.

ABSTRACT

Purpose: In general, articles on diagnostic tests have a very poor methodological quality. If we translate their conclusions to daily practice without a proper analysis it is easy to see a typical trend: physicians use new (and expensive) tests without increasing diagnostic capacity; they only increase the health budget.

Evidence-based medicine (EBM) consists of using the best evidence in decision-making. It deals with selected and relevant information, supported by data obtained through the most rigorous scientific method: epidemiology and statistics.

Critical evaluation of papers related to diagnostic tests is our aim. We provide with basic skills for evaluation and analysis of papers by means of simple topics on statistics, design of clinical assays and clinical epidemiology.

Methods: Review of the medical literature.

Results: To evaluate papers on diagnostic tests and to use a test correctly, it is necessary to know its diagnostic capacity, the level of certainty to start treatment, the probability of having the disease before using the test and the test capacity to change that probability.

Recibido: 8/11/07. Aceptado: 20/10/08.

Servicio de Oftalmología. Hospital Universitario Bellvitge. Barcelona. España.

[1] Doctor en Medicina.

[2] Licenciado en Medicina.

Correspondencia:
J.C. Mesa Gutiérrez
Servicio de Oftalmología. Hospital Universitario de Bellvitge
Feixa Llarga, s/n
08907 Hospitalet de Llobregat (Barcelona)
España
E-mail: jcarlosmesa@mixmail.com

Resultados: Para evaluar artículos sobre pruebas diagnósticas y para utilizar eficientemente una prueba diagnóstica necesitamos conocer su capacidad diagnóstica, el nivel de certeza que queremos tener para llevar a cabo una actitud terapéutica, la probabilidad de que el paciente tenga la enfermedad antes de utilizar la prueba y la capacidad que dicha prueba tiene de modificar tal probabilidad.
Conclusiones: El rendimiento máximo de una prueba se obtiene cuando nos encontramos en una situación de máxima incertidumbre (probabilidad de enfermedad del 50%). Su mayor utilidad se obtiene cuando aumentamos al máximo la probabilidad preprueba, tras una minuciosa anamnesis y exploración.

Palabras clave: Medicina Basada en la Evidencia, razón de verosimilitud, probabilidad, odds, valor predictivo, sensibilidad, especificidad, teorema de Bayes.

Conclusions: The maximum performance of a test is reached when it is used in a maximum uncertainty situation (disease probability 50%). Its highest usefulness is obtained when pre-test probability is maximal, after a careful review of the patient's history and a complete patient examination *(Arch Soc Esp Oftalmol 2008; 83: 639-652).*

Key words: Evidence-Based Medicine, likelihood ratios, probability, odds, predictive value, sensitivity, specificity, Bayes' theorem.

INTRODUCCIÓN

La medicina basada en la evidencia (MBE) es el uso consciente, explícito y juicioso de la mejor evidencia disponible en la toma de decisiones sobre los cuidados de pacientes individuales (1). Se trata de actuar en la clínica utilizando información seleccionada y relevante, y que venga avalada por datos obtenidos a través del método científico más riguroso: la epidemiología y la estadística. La MBE se centra en el examen riguroso de las pruebas científicas suministradas por la investigación clínica.

La práctica médica tradicional, heredera de los paradigmas clásicos, pero vigente todavía hoy en muchos ámbitos, se puede caracterizar por la creencia en que:

a) Las observaciones derivadas de la experiencia clínica personal son una forma válida de generar, validar y transmitir los conocimientos acerca del pronóstico de las enfermedades, el rendimiento clínico de las pruebas diagnósticas y la eficacia o eficiencia de los tratamientos.

b) Un buen conocimiento de la teoría fisiopatológica subyacente, una combinación de habilidad en el razonamiento y la especulación lógica, y una buena dosis de sentido común permiten interpretar adecuadamente los signos de la enfermedad y elegir el tratamiento más adecuado; y

c) Los conocimientos se actualizan adecuadamente mediante libros de texto y revistas con los tradicionales «artículos de revisión», en los que los expertos de más experiencia y prestigio nos seducen con sus opiniones juiciosas acerca de las soluciones a los problemas clínicos. Es por ello que se concede una enorme credibilidad al argumento de autoridad, y los apartados «introducción» y «discusión» de los trabajos de investigación original centran los debates y marcan las pautas de actuación para una «buena praxis».

Frente al modelo tradicional, la MBE se caracteriza por la creencia en que:

a) La información derivada de la experiencia clínica y la intuición puede llevar a conclusiones erróneas si no se basa sólidamente en observaciones sistemáticas;

b) El estudio y conocimiento de los mecanismos teóricos básicos de la enfermedad es necesario pero insuficiente para guiar la práctica clínica; y

c) El profesional necesita conocer ciertas reglas para evaluar rigurosamente la metodología con la que se han obtenido las pruebas científicas en las que se sustentan sus decisiones.

Además, el sistema tradicional de reciclaje de conocimientos, mediante sesiones de «Formación Médica Continuada», ha quedado obsoleto, puesto que los libros de texto tradicionales son incapaces de recoger la nueva información científica que se produce en tiempo real. Así, después de 43 ensayos clínicos aleatorizados (en los que participaron más de 21.000 pacientes) en que se demostraba la efica-

cia de la trombólisis temprana sobre la mortalidad del infarto de miocardio, ningún libro de texto médico establecía aún esta indicación como rutinaria. Pero aún hay más: en 1990 y después de 15 ensayos clínicos aleatorizados y tres metaanálisis, se seguía recomendando en los textos especializados la administración profiláctica de lidocaína para prevenir el re-infarto, un medida completamente ineficaz (2).

Por otra parte, en las últimas décadas hemos asistido al fenómeno de la explosión informativa, un crecimiento exponencial de la literatura médica que hace materialmente imposible mantenerse al día si pretendemos hacerlo utilizando este recurso de una manera acrítica. En 1948 había cerca de 4.700 revistas científicas; en 1994 se publicaban unos 2 millones de artículos en 20.000 revistas médicas (3). Aunque quizá en nuestra especialidad sea más sencillo dado (por el momento) el menor volumen de información a consumir, se estimó que para mantenerse al día, un generalista debería leer 19 artículos al día durante los 365 días del año (4). Si combinamos este hecho con la mejora actual en el acceso a la información gracias internet, el resultado es una avalancha de información que obliga al médico que no quiera ir perdiendo competencia profesional con el paso del tiempo, a dominar habilidades y técnicas sistemáticas que le doten de sentido crítico y le permitan identificar la información verdaderamente relevante para su práctica «a pie de paciente». Necesitamos el imán que nos permita buscar la aguja en el pajar, puesto que con el actual ritmo de producción de ensayos clínicos y otras investigaciones rigurosas, la cuestión ha dejado de ser si nuestras actuaciones en la práctica tienen buena base científica, sino cuánta de la evidencia actualmente disponible se aplica en la práctica diaria (5).

La MBE resta fuerza a la intuición, la experiencia clínica no sistematizada y la fisiopatología como elementos suficientes para la toma de decisiones clínicas, y acentúa el valor del examen riguroso de las pruebas científicas suministradas por la investigación clínica. Para ello incorpora al arsenal de saberes y habilidades básicas para el desempeño de la profesión médica, la destreza en el uso de una serie de conocimientos sencillos de estadística, diseño de investigaciones y epidemiología clínica.

Los médicos deben, pues, adquirir la responsabilidad de evaluar de forma crítica e independiente la credibilidad tanto de las evidencias como de las opiniones ofertadas. Lo importante no es el mensaje, sino el método con el que se ha llegado a los datos. Son ahora los apartados «material y métodos» y «resultados» de los artículos de investigación los que se convierten en las piezas claves de los trabajos médicos, pues son las secciones que deben evaluarse detalladamente para valorar la validez de los datos que aportan. Así que para este nuevo estilo científico de hacer medicina, la autoridad establecida –los «expertos»–, tienen mucho menor peso. Por eso Sackett (6), convertido a su vez en experto muy a su pesar, aboga por la desaparición de esta figura para facilitar el avance de la ciencia: primero por la tendencia existente en el resto de la comunidad médica a no contradecirlos, ya sea por deferencia, miedo o respeto; segundo, porque los editores de las revistas se enfrentan a la tentación de aceptar o rechazar nuevas ideas y evidencias en función de su coincidencia o no con la opinión «experta».

Esto no debe interpretarse como un rechazo a lo que uno puede aprender de sus maestros o colegas. Únicamente significa que, si se busca la mejor atención para nuestros enfermos, una «buena praxis» de la medicina moderna debe necesariamente partir de un conocimiento riguroso de las pruebas científicas que sustentan cada una de sus prácticas clínicas.

Suministraremos las herramientas necesarias para la evaluación de la calidad de la información proporcionada por artículos que tratan sobre pruebas diagnósticas.

SUJETOS, MATERIAL Y MÉTODO

Se ha realizado una revisión bibliográfica de la información disponible que aborda el tema planteado. Al ser la Medicina Basada en Evidencias una disciplina relativamente reciente (el término fue acuñado en los años 80 por un grupo de epidemiólogos clínicos canadienses de la Universidad de McMaster y su difusión en la práctica clínica se produjo a partir de 1992) su sistemática de trabajo se divulga fundamentalmente en Internet y por tanto, la información disponible se encuentra de forma mayoritaria, y completamente gratuita, en la red. Así, en este trabajo casi toda la bibliografía, las hojas de cálculo, y las herramientas utilizadas se han obtenido de las numerosas sites y webs disponibles en la red sobre este nuevo «estilo de proceder médico» que va arraigándose progresivamente en la comunidad médica.

Analizada esta información, se seleccionó un artículo sobre pruebas diagnósticas publicado en

una revista de impacto (Archives of Ophthalmology) y se procedió a su análisis crítico a modo de ejemplo.

RESULTADOS

Así pues, se trata de actuar en la clínica utilizando información seleccionada y relevante, y que venga avalada por datos obtenidos a través del método científico más riguroso: la epidemiología y la estadística. Ello no significa que haya que ser un entendido en epidemiología o en estadística para aplicar los principios de la MBE: es factible adquirir unas habilidades básicas que nos posibiliten tener juicio crítico para obtener la mejor evidencia científica del tema que nos interese (7).

La búsqueda de información

Las fuentes para responder a nuestra pregunta clínica son varias. Podemos recurrir a libros de texto tradicionales pero, como hemos visto, la información que contienen queda obsoleta con rapidez y no son adecuados para responder a preguntas de 3 componentes.

Se puede recurrir a bases de datos con filtro de calidad, como Embase, base de datos del repertorio Excerpta Medica; o la popular *Medline* (www.ncbi.nlm.nih.gov), base de datos del repertorio Index Medicus, producido por la *National Library of Medicine* y de libre acceso, a diferencia de Embase, gracias a la administración Clinton. La propia base de datos Medline posee un buscador de preguntas clínicas *(Clinical Queries)* donde se pueden introducir los términos de búsqueda para terapia, diagnóstico, pronóstico o etiología, facilitando enormemente la realización de este tipo de búsquedas.

A partir de esta base de datos obtendríamos los artículos de interés, pero éstos deberán ser evaluados para valorar la evidencia que aportan. Una tercera alternativa es la búsqueda en revistas secundarias, que realizan el proceso evaluador por nosotros y nos ofrecen información ya revisada y catalogada desde el punto de vista de la evidencia a partir de artículos metodológicamente sólidos. Hoy en día existen en la red diferentes fuentes que proporcionan información de este tipo. En castellano se puede acceder a ellas a través de páginas como www.fisterra.com, o www.infodoctor.org, La página web de la Universidad de Washington ofrece vínculos a distintas fuentes de MBE, incluyendo la colaboración Cochrane (http://healthlinks.washington.edu/ebp/ebpresources.html).

El proceso diagnóstico

Para leer críticamente un artículo sobre pruebas diagnósticas y para aprender a usar de manera eficiente una prueba diagnóstica, debemos entender previamente una serie de requisitos básicos:

1. Una prueba diagnóstica es útil, desde el punto de vista clínico, sólo si nos induce a tomar las decisiones (terapéuticas) adecuadas.
2. El diagnóstico es una actividad que el médico sólo debe desarrollar en un ambiente de incertidumbre. Es decir, que el uso de los tests diagnósticos tiene sentido únicamente cuando la anamnesis, la exploración física y otras pruebas de diagnóstico básico no nos han proporcionado la suficiente certeza como para llevar a cabo una actitud terapéutica, y hablamos de «suficiente» certeza porque no necesitamos una certeza absoluta para iniciar un tratamiento.
3. Sólo tiene sentido plantear el uso de nuevos tests si sabemos que sus resultados van a ser capaces de disminuir nuestra situación de incertidumbre (fig. 1).

Así, el uso «racional» de un test requiere que el clínico:

a) Conozca la probabilidad de que el paciente presente la enfermedad antes de hacer el test (probabilidad *a priori* o preprueba);

b) Conozca la capacidad que tiene el test de modificar esa probabilidad (probabilidad *a posteriori* o postprueba), y

c) Establezca el nivel de certeza que necesita tener para tomar una decisión terapéutica (umbral de acción).

El grado de incertidumbre/certeza que tenemos de que ocurra un evento (p. ej., que el paciente que

Fig. 1: Umbral de decisión.

tenemos delante tenga, finalmente, una determinada enfermedad), puede expresarse de dos maneras equivalentes (fig. 2): mediante una probabilidad o mediante una «odds». Una probabilidad (un riesgo) es una cantidad entre 0 y 1 que coincide con la frecuencia de aparición del evento, expresada como el número de casos favorables partido por el total de casos. La «odds», usada ampliamente por sus ventajas para el cálculo, es el mismo concepto pero expresado de una manera menos intuitiva: con una cantidad que oscila entre 0 e ∞ calculada con el número casos favorables partido por el de desfavorables (una probabilidad dividida por su complementaria).

Para conocer la capacidad que tiene el test diagnóstico que vamos a solicitar de cambiar esa incertidumbre, utilizamos la sensibilidad, la especificidad y los valores predictivos. La sensibilidad y la especificidad se consideran los parámetros que mejor evalúan el rendimiento diagnóstico (validez interna) de una prueba: la sensibilidad (S) representa la capacidad que tiene el test para detectar a los casos, y la especificidad (E), la capacidad que tiene el test para detectar a los sanos (no casos). Matemáticamente ambas son probabilidades condicionales, y se expresarían de la siguiente forma: Sensibilidad = p(+/E), léase como *probabilidad de que el test sea positivo dado que el sujeto está enfermo;* Especificidad = p(-/noE), *probabilidad de que el test sea negativo dado que el sujeto está sano (no enfermo).* Ambos valores se obtienen tras aplicar la prueba a poblaciones en las que se conoce con certeza su estatus de enfermedad (fig. 3). Una prueba extremadamente sensible (S) se utiliza para descartar la presencia de enfermedad, y se define como «SnOUT», regla mnemotécnica que indica que un resultado negativo (n) descarta completamente (OUT) la presencia de enfermedad. Por el contrario, una prueba extremadamente específica (ES) se utiliza para asegurar la presencia de enfermedad, y se define como

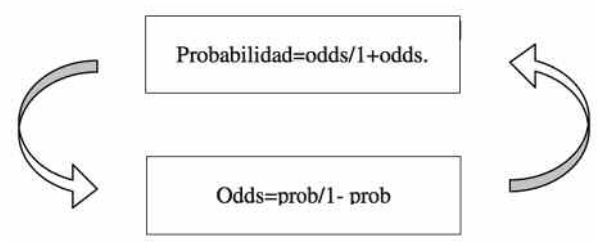

Fig. 2: Relación entre odds y probabilidad.

«ESpIN», regla mnemotécnica que indica que un resultado positivo (p) diagnostica (IN) la enfermedad. Aunque idealmente las pruebas diagnósticas deberían tener una alta sensibilidad y especificidad, por regla general ambos parámetros guardan una relación inversa, que viene representada por la llamada curva ROC («receiver-operating characteristic») (8).

Pero ¡ojo!, la dificultad capital consiste en que en la práctica clínica diaria no necesitamos estos parámetros: la incertidumbre radica exactamente en que desconocemos el estado de salud del sujeto que tenemos delante, y lo que conocemos con certeza es el resultado del test. La pregunta que nos hacemos a pie de paciente es si el resultado positivo o negativo de la prueba es correcto o no. La respuesta son otras probabilidades condicionales: los valores predictivos del test. El valor predictivo positivo (VPP) es p(E/+), *probabilidad de que el sujeto esté enfermo dado que el test es positivo* (un signo patognomónico tendrá un VPP del 100%), y el negativo (VPN) es p(noE/-), *probabilidad de que el sujeto esté sano dado que el test es negativo.* Observe cómo la expresión p(E/+) (valor predictivo positivo) es absolutamente diferente de p(+/E) (Sensibilidad). No es una sutileza sin importancia, sino la conocida «falacia de transposición de los condicionales». El valor predictivo global se define como la probabilidad que tiene una prueba de acertar (fig. 3).

¿Qué problema plantea el uso de los valores predictivos? Pues que su cálculo no es directo desde la sensibilidad y la especificidad, sino que dependen de la prevalencia de enfermedad: la probabilidad de enfermedad previa a hacer el test. Las probabilidades condicionales se rigen mediante el teorema de Bayes (fig. 4), y si expresamos el valor predictivo a partir de este teorema comprobaremos que el resultado final depende de la prevalencia. Esto explica porqué los resultados de las pruebas diagnósticas varían al aplicarlos de una región a otra cuando la prevalencia de enfermedad en ambas zonas es muy diferente.

Por último, para usar eficientemente una prueba diagnóstica necesitamos el nivel de certeza que queremos tener para llevar a cabo una actitud terapéutica.

Una vez aclarados estos conceptos nos centraremos en la lectura crítica. Los requisitos para valorar críticamente los artículos sobre pruebas diagnósticas se exponen en la tabla I y equivalen a las preguntas que debe responder el artículo en cuestión: ¿son los resultados válidos?, ¿cuáles son?, ¿me serán útiles? (9)

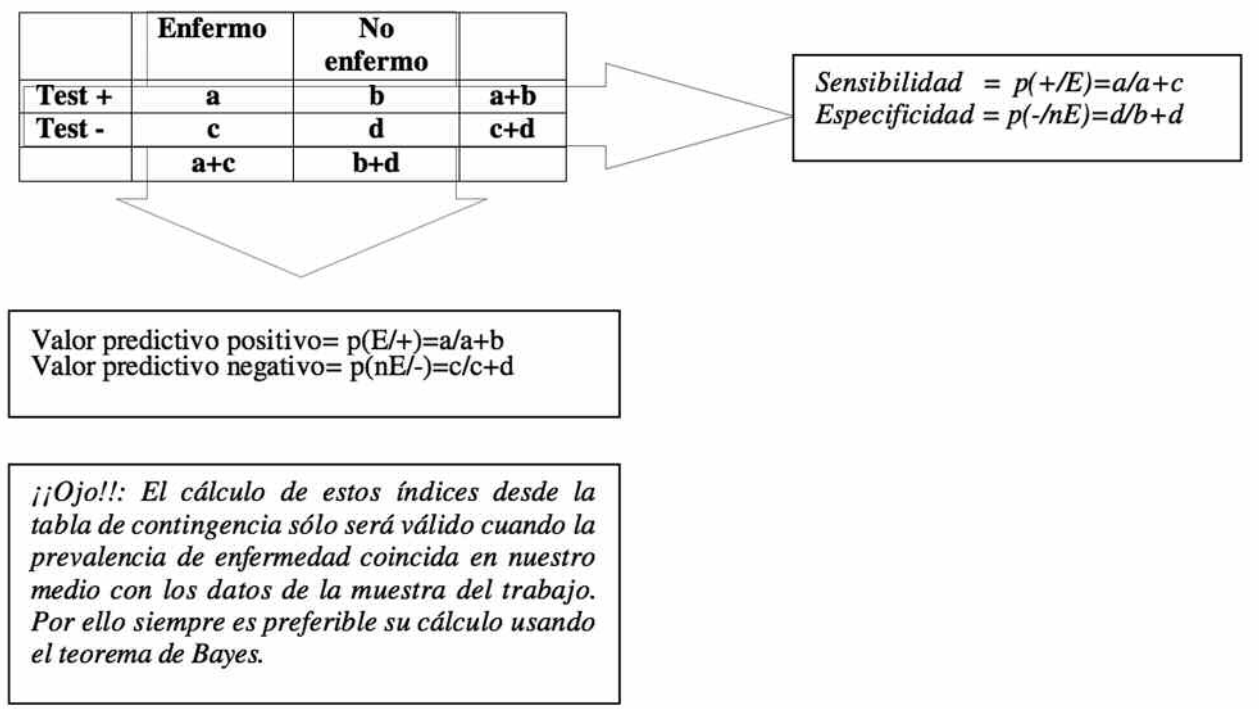

Fig. 3: Determinación de la sensibilidad, la especificidad y los valores predictivos a partir de una tabla 2x2.

1. ¿Son válidos los resultados?

a) Comparación independiente (y ciega) con una prueba de referencia

Utilizaremos el artículo «Uso del cambio glaucomatoso progresivo del disco óptico como referencia estándar en la evaluación de las pruebas diagnósticas de glaucoma» publicado en 2005 en la revista *American Journal of Ophthalmology* (10).

Toda prueba diagnóstica ha de ser comparada con la «verdad», con una prueba lo más objetiva posible que diagnostique de forma certera la enfermedad. Es el denominado «patrón oro» o prueba de referencia (p. ej., la anatomía patológica para un tumor o la coronariografía para la enfermedad coronaria). Si la prueba de referencia elegida es poco válida, los resultados del estudio pueden ponerse en entredicho. Cuando la prueba de referencia es invasiva, en los casos dudosos suele optarse por un seguimiento del paciente durante un tiempo razonable para confirmar/descartar la enfermedad, ya que no se justificaría la realización de una prueba invasiva y con riesgos añadidos en ausencia de signos de enfermedad.

Además de contar con una prueba de referencia adecuada, ésta debe aplicarse independientemente del resultado de la prueba en estudio. En caso de que el patrón oro se aplique sólo cuando la prueba en estudio detecta la enfermedad se produciría el llamado *sesgo de verificación diferencial (work-up detection bias).* Un ejemplo sería la realización de una coronariografía sólo cuando el test de esfuerzo es positivo. En los estudios prospectivos este sesgo

Tabla I. Preguntas genéricas para un artículo sobre pruebas diagnósticas

¿Son válidos los resultados del estudio?
¿Existe una comparación independiente con una prueba de referencia adecuada en todos los casos?
¿El espectro de pacientes de la muestra es adecuado? (similar a los que en la práctica clínica se aplicará el examen diagnóstico)
¿Se describen los métodos con el suficiente detalle como para ser reproducible?
¿Cuáles son los resultados?
¿Los datos del estudio permiten calcular los cocientes de probabilidad (likelihood ratios)?
¿Cuál es la precisión de los resultados?
Aplicabilidad de los resultados
En su escenario de trabajo ¿es factible la reproducibilidad de la prueba y su interpretación?
En caso de que así sea, ¿es aceptable? A considerar disponibilidad, costes, riesgo/beneficio
¿Modificarán los resultados de la prueba la decisión sobre cómo actuar?

Teorema de Bayes
P(A/B)= p(A)*p(B/A)/p(A)*p(B/A)+p(nA)*p(B/nA)

Teorema de Bayes aplicado al valor predictivo positivo (VPP)
P(E/+)=p(E)*p(+/ E)/p(E)*p(+/E)+p(nE)*p(+/ nE)

Donde
p(E) = Probabilidad de estar enfermo = Prevalencia
p (+/E) = Probabilidad de test positivo si está enfermo = Sensibilidad
p(nE) = Probabilidad de no estar enfermo = 1- Prevalencia
p(+/nE) = Probabilidad de test positivo si no está enfermo (Falso positivo) =1- Especificidad

Así pues, tendríamos:
VPP= Prevalencia*Sensibilidad/ Prevalencia*Sensibilidad+ (1-Prevalencia)*(1-Especificidad)

Teorema de Bayes aplicado al valor predictivo negativo (VPN)
P(noE/-)= p(noE)*p(-/noE)/p(E)*p(-/E)+p(noE)*p(-/noE)

Donde
p(noE) = Probabilidad de estar sano = 1- Prevalencia
p(-/ noE) = Probabilidad de test negativo si está sano = Especificidad
p(E) = Probabilidad de estar enfermo = Prevalencia
p(-/E) = Probabilidad de test negativo si está enfermo (Falso negativo) =1- Sensibilidad

Así pues, tendríamos:
VPN =(1-Prevalencia)* Especificidad/Prevalencia*(1-Sensibilidad)+(1-Prevalencia)*Especificidad.

Fig. 4: Teorema de Bayes y su aplicación a la determinación del valor predictivo.

puede resolverse con un seguimiento adecuado de los pacientes que presentan un resultado negativo, mientras que en los retrospectivos no es posible eliminar el sesgo.

De forma ideal, la interpretación de la prueba debe realizarse de forma ciega, es decir, que el patrón oro debe aplicarse sin conocer el resultado del test, y el test sin conocer los resultados del patrón oro, pero esto no es siempre posible. Así, en los estudios sobre pruebas de imagen y glaucoma es posible que el oftalmólogo conozca datos clínicos del paciente, y tome su decisión conociendo el resultado de la polarimetría láser (GDx), lo que introduce un sesgo en la estimación del rendimiento del test. Además, para las pruebas de imagen hay que considerar el posible *sesgo del observador,* por el que las expectativas del evaluador influyen en el resultado de la medición. Si el oftalmólogo tiene información clínica sobre el paciente, dicha información influirá en la interpretación de las imágenes. De ahí la importancia de que la interpretación se realice de forma ciega. En el artículo que nos ocupa, los datos se recogieron independientemente de la evaluación del oftalmólogo y antes de que se obtuvieran las pruebas de imagen.

Los errores sistemáticos o sesgos son errores producidos en el diseño o ejecución del estudio y hace que los resultados de la muestra sean diferentes de la población de la que proceden. No se correlacionan con el tamaño de la muestra y cuando no se controlan tienden a invalidar las condiciones del estudio, conduciendo a la elaboración de conclusiones incorrectas. En los artículos sobre pruebas diagnósticas interesa el sesgo diagnóstico o de Berkson: para saber qué ocurre en la población se elige una muestra hospitalaria de esa población y el factor de riesgo que estudiamos se asocia a una mayor probabilidad de hospitalización. También cometemos este sesgo cuando elegimos como control pacientes con alguna enfermedad que también se asocia al factor de exposición estudiado.

b) Espectro adecuado de pacientes

Los pacientes incluidos en el estudio deben constituir una población muy similar a la que habitualmente se solicitaría el estudio en la práctica cotidiana. Un error en la selección de los pacientes es comparar sujetos claramente enfermos, en los que no existe duda diagnóstica, con sujetos sanos, lo que da lugar a resultados decepcionantes cuando la prueba se aplica en la práctica clínica real.

Un problema similar se produce si la proporción de pacientes con procesos avanzados es superior a lo habitual, produciendo un aumento de la sensibilidad. En la mayoría de artículos sobre pruebas diagnósticas relacionadas con el glaucoma el diagnóstico se realiza por la presencia de anomalías estructurales importantes o de alteraciones en el campo visual en la perimetría estándar automatizada. Sin embargo, la utilización de estos parámetros como estándar de referencia para el diagnóstico de glaucoma tiene una gran limitación: la alteración estructural de la papila o la anomalía en el campo visual de un paciente glaucomatoso sólo puede objetivarse después de que se hayan perdido un número considerable de fibras nerviosas. Por tanto, los pacientes incluidos en estos estudios tendrán probablemente un estadio más avanzado de la

enfermedad, lo que puede facilitar su identificación por la prueba de imagen. Es muy probable que en aquellos casos muy evolucionados el rendimiento aumente; sin embargo es muy infrecuente que el artículo especifique el porcentaje de casos incipientes y el de evolucionados.

Otro problema a considerar es el de las pérdidas. Es posible que en algunos sujetos no se pueda aplicar el «gold standard » o el test en cuestión, o que, si se aplica, el resultado no sea concluyente. En el artículo analizado los ojos con errores refractivos importantes se excluyeron del estudio porque las pruebas de imagen tienden a ser menos fiables en estos pacientes.

Una cuestión más: si el diseño del estudio es tipo caso/control, la prevalencia de enfermedad en la muestra no será igual a la de la población habitual, puesto que el autor selecciona uno, dos o «n» sujetos sanos por cada enfermo. En consecuencia, los valores predictivos del test no serán aplicables a su muestra habitual, a no ser que la selección de controles se haya realizado de tal forma que mantenga la misma prevalencia que luego encontraría en la práctica. ¿Cuál fue la muestra de pacientes en el estudio que estamos analizando? Se trata de un estudio caso-control observacional. Fueron pacientes incluidos en otro estudio, el estudio DIGS *(Diagnostic Innovations in Glaucoma Study)* diseñado para evaluar la estructura del nervio óptico y la función visual en el glaucoma. Los pacientes seleccionados debían cumplir unos criterios de inclusión, como es la constatación de cambio estructural, y tener un cambio glaucomatoso progresivo documentado mediante fotografías estereoscópicas basándose en el adelgazamiento del anillo neuro-retiniano, el aumento de la excavación y el aumento del defecto de la capa de fibras nerviosas. Las diferencias en el color del anillo neuro-retiniano, la presencia de hemorragias en la papila a la atrofia papilar no fueron suficientes para diagnosticar la progresión glaucomatosa. En principio, y con estos datos, la población podría no asemejarse a la que podemos encontrar en nuestro medio.

c) Descripción de los métodos utilizados

La descripción del test debe tener el suficiente detalle para que la prueba sea aplicable en sus pacientes y debe incluir tanto la preparación del paciente como la realización de la prueba y la interpretación de los resultados. El artículo habla del GDx con compensación corneal variable (GDx-VCC).

Otra información que debiera incluir el estudio pero que generalmente no se ofrece son los datos sobre la reproducibilidad del test, sobre todo si se trata de pruebas de imagen o de valoraciones «subjetivas». Si el resultado del test cambia según el observador, el test no será fiable. Únicamente citaremos en este punto que cuando se trata de variables categóricas la concordancia entre observadores suele medirse mediante el índice kappa, y en caso de variables continuas mediante métodos gráficos como el de Bland-Altman (11). En el artículo que estamos revisando no se ofrece ninguna información de este tipo y, aunque la valoración de los parámetros del GDx no plantearía problema alguno, la valoración clínica de la papila conlleva cierto grado de subjetividad: el cambio progresivo en la papila se evaluó basándose en el adelgazamiento del anillo neuro-retiniano, el aumento de la excavación y el aumento del defecto de la capa de fibras nerviosas.

Sin embargo, se ha sugerido que las características estructurales de la papila no deberían utilizarse como criterio de inclusión en los estudios que evalúan la precisión diagnóstica de los pruebas de imagen. Ya que estos instrumentos evalúan los hallazgos estructurales de la papila o de la capa de fibras nerviosas, la inclusión de sujetos basados en anomalías en estas estructuras conduciría a una sobrestimación de la sensibilidad del instrumento.

Otra forma de realizar una lectura crítica es mediante las tablas del acuerdo STARD *(STAndards for the Reporting of Diagnostic accuracy studies),* cuya finalidad es mejorar la fiabilidad y exactitud de los artículos sobre diagnóstico, de forma que permita a los lectores estimar la existencia de sesgos (validez interna) y evaluar si sus conclusiones son generalizables (validez externa). Se trata de una lista de verificación con 25 ítems que el lector debe comprobar dentro de los distintos apartados del artículo (título y palabras clave, métodos, resultados y discusión). Dicha tabla puede consultarse y descargarse de forma gratuita en su página web: www.stard-agreement.org (12).

2. ¿Cuáles son los resultados?

El artículo debe incluir los cocientes de verosimilitud *(likelihood ratios)* o por lo menos debe proporcionar los datos necesarios para su cálculo. Quizás

este término resulte menos familiar en comparación con la sensibilidad, la especificidad o los valores predictivos, pero comprobaremos que tiene más ventajas. El artículo nos da los resultados como curvas ROC, cuya interpretación clínica suele ser difícil. El cálculo de la LR proporciona una información más fácil de comprender.

Toda la información que necesitamos sobre la habilidad de una prueba diagnóstica para diagnosticar/descartar la enfermedad con independencia de la prevalencia la tenemos en el *cociente de verosimilitud, razón de verosimilitud (RV) o likelihood ratio (LR).* Este término puede encontrase en muchos textos en castellano como cociente o razón de probabilidades, pero en sentido estricto una función de probabilidad tiene propiedades matemáticas bien diferenciadas de una razón de verosimilitud (13). Sin embargo, esta definición es la que se ha popularizado al ser la elegida por el grupo español de MBE. El cociente de verosimilitud del resultado de un test se define como la probabilidad de ese resultado en enfermos dividida por la probabilidad de ese mismo resultado en no enfermos. Así definida, la LR constituye la evidencia que proporciona cada uno de los resultados del test a favor (o en contra) de la enfermedad. Indica cuánto aumenta o disminuye la probabilidad de enfermedad. Funciona como el riesgo relativo (RR), ya que nos indica cuántas veces es más probable ese resultado en los enfermos frente a los no enfermos (14). Una LR=1 nos indicaría que el resultado es igual de probable en ambos y, por lo tanto, no diferencia sanos de enfermos, al igual que un RR=1 nos indica que el riesgo de muerte es igual en tratados que en no tratados. Aunque su definición parezca complicada su cálculo es bien sencillo a partir de la sensibilidad y la especificidad (fig. 5).

¿Cómo me ayuda la LR en la práctica? Al fin y al cabo, saber que el resultado de un test es, por ejemplo, dos veces más frecuente en un enfermo que en un sano no es tan informativo como saber la probabilidad que tiene el sujeto de estar enfermo, cosa

***Likelihood* ratio para test positivo (LR+) = Sensibilidad/1-Especificidad.**

***Likelihood* ratio para test negativo (LR-) = 1-Sensibilidad/Especificidad.**

Odds pre-test*LR=Odds post-test.

Fig. 5: Cálculo de la razón de verosimilitud (likelihood ratio) y su relación con la odds.

que sí hace el valor predictivo (p. ej., 85%). Veamos cómo utilizando la LR también podemos obtener esa probabilidad: en la práctica, cuando sospechamos una enfermedad lo que hacemos es solicitar una prueba diagnóstica que nos permita confirmar o descartar nuestra sospecha. El resultado del test modifica la probabilidad previa que nosotros estimábamos para ese paciente, que es lo que se denomina también probabilidad pre-test. Un ejemplo: ¿Cuál es la probabilidad a priori de padecer glaucoma? De forma genérica, el riesgo a lo largo de nuestra vida estaría en torno al 0,5%, que equivaldría a su incidencia; la prevalencia es del 1% (15).

Entre los pacientes que son evaluados por aumento de la presión intraocular, alteraciones en la papila y anomalías en al campo visual la probabilidad aumenta. Es decir: cada dato de la exploración física y de las exploraciones complementarias practicadas va alterando la probabilidad de enfermedad.

Así pues, en la práctica lo que hace la LR es modificar la probabilidad que nosotros habíamos estimado (probabilidad preprueba), dando un nuevo valor que es la probabilidad postprueba. En realidad este cambio de probabilidades no puede hacerse de forma directa, puesto que lo que modifica la LR no es la probabilidad sino la *odds.* Por tanto habría que transformar la probabilidad preprueba en *odds,* multiplicarla por la LR y volver a transformar el resultado (odds postprueba) en probabilidad. Si le parece complicado realizar las traslaciones entre odds y probabilidades puede optar por el nomograma de Fagan (16) (fig. 6). Como puede apreciarse, en la primera columna se presentan las probabilidades pre-test posibles, en la segunda los posibles valores de RV y en la tercera las probabilidades postest. Para conocer la probabilidad de nuestro paciente de tener una determinada enfermedad sólo hay que trazar una línea que una las dos primeras columnas según los valores adecuados y obtendremos, al prolongarla, un valor en la tercera columna que no es más que la probabilidad postest.

Cuanto más extremo sea el valor de la LR más importante es el rendimiento diagnóstico de la prueba. En clínica, hay que usar test cuya LR para positivos sea muy alta y/o cuya LR para negativos sea muy baja. En la tabla II se presenta su interpretación en función de su valor. La precisión de los resultados debe expresarse con el intervalo de confianza al 95% de la LR de cada resultado del test.

Los autores del artículo seleccionado nos ofrecen como resultado únicamente la sensibilidad, las áre-

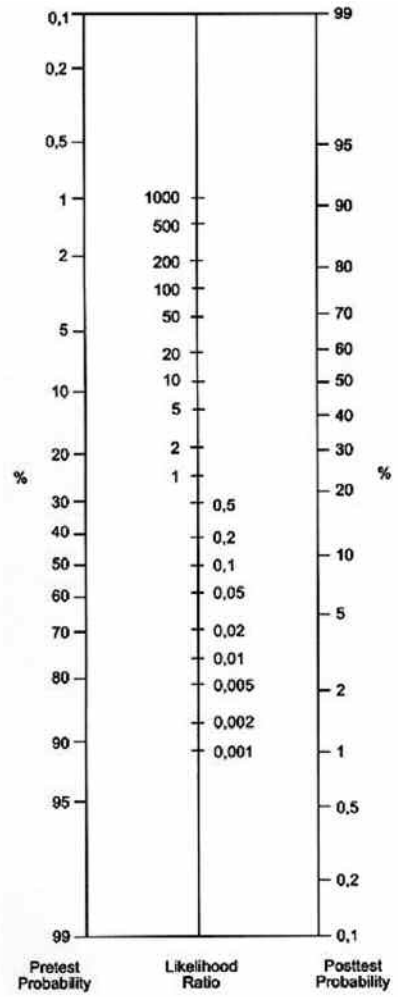

Fig. 6: Nomograma de Fagan.

as bajo las curvas ROC, el análisis de la varianza y la LR para un resultado positivo (recuerde que el objetivo del trabajo era detectar aquellos casos con alta probabilidad de glaucoma). El estudio incluye finalmente 284 sujetos: 61 sujetos sanos (verdaderos negativos) y 71 enfermos (verdaderos positivos). A partir de estos datos es posible calcular el resto de parámetros con las fórmulas vistas, pero una opción más sencilla es introducir los datos proporcionados por el estudio en la calculadora que ofrece la red CASPe en su página web (16), que no es más que una hoja de cálculo. La tabla III muestra esos datos introducidos en la tabla 2x2 que proporciona la calculadora y la tabla IV muestra el resultado final. Compruebe que la hoja de cálculo nos ofrece todos los valores que nos interesan y también que los resultados coinciden con los presentados por los autores.

¿Qué podemos decir del resultado del test? Su LR para resultados positivos es 1,63 o, lo que es lo mismo, es un test prácticamente inútil para el diagnóstico de glaucoma (en su uso como prueba diagnóstica inicial y en nuestro escenario). Y su LR para resultados negativos es 0,83: inútil para descartar glaucoma (con las mismas consideraciones anteriores).

En cuanto a la precisión de los resultados, no se preocupe, ya que la hoja de cálculo le ofrece el intervalo de confianza al 95%.

Tabla II. Interpretación clínica de la razón de verosimilitud (*likelihood* ratio)

LR+ > 10: Aumento grande: Test excelente
LR+ entre 5 y 10: Aumento moderado: Test bueno
LR+ entre 5 y 2: Aumento pequeño: Test malo
LR+ < 2: Aumento insignificante: Test inútil
LR=1: Sin cambios
LR- entre 0,5 y 1: Descenso insignificante: Test inútil
LR- entre 0,2 y 0,5: Descenso pequeño: Test malo
LR- entre 0,1 y 0,2: Descenso moderado: Test bueno
LR- < 0,1: Descenso grande: Test excelente

DISCUSIÓN

1) ¿Modificarán los resultados de la prueba la decisión sobre cómo actuar?

Una prueba diagnóstica es útil desde el punto de vista clínico sólo si nos induce a tomar las decisiones (terapéuticas) adecuadas en un ambiente de incertidumbre.

Básicamente, debemos realizar tres pasos sucesivos una vez estudiado el escenario: primero determinar el umbral de acción (UA); segundo, determinar –en las condiciones estrictas del paciente del escenario– cuál es la probabilidad de enfermedad antes de hacer la prueba y, por último, calcular la

Tabla III. Tabla de contingencia de la calculadora CASPe

	Característica evaluada Presente (Prueba de referencia +)	Ausente (Prueba de referencia -)	
Prueba diagnóstica +	71	18	89
Prueba diagnóstica -	130	65	195
Total	201	83	284

Tabla IV. Resultado ofrecido por la calculadora CASPe a partir de la tabla III

		IC 95%	
Sensibilidad	35,3%	29,0%	a 42,2%
Especificidad	78,3%	68,3%	a 85,8%
Valor predictivo positivo	79,8%	70,3%	a 86,8%
Valor predictivo negativo	33,3%	27,1%	a 40,2%
Proporción de falsos positivos	21,7%	14,2%	a 31,7%
Proporción de falsos negativos	64,7%	57,8%	a 71,0%
Exactitud	47,9%	42,1%	a 53,7%
Odds ratio diagnóstica	1,97	1,09	a 3,58
Índice J de Youden	0,1		
CPP o LR(+)	1,63	1,04	a 2,55
CPN o LR(-)	0,83	0,71	a 0,96
Probabilidad pre-prueba (Prevalencia)	70,8%		

probabilidad posprueba de enfermedad dados los resultados del test. Si el resultado del test es capaz de llevar la probabilidad de enfermedad más allá del valor umbral, el uso del test está clínicamente justificado (en las condiciones del escenario).

2) Determinar el umbral de acción

Recordemos el escenario que plantea el artículo analizado: pretende establecer el valor de una prueba de imagen (GDx) en el glaucoma.

Veamos cómo nuestras herramientas nos ofrecen una solución razonable. En la situación del escenario, lo primero que debemos hacer es determinar el UA: qué probabilidad es suficiente para indicar el tratamiento. Es decir, ¿a partir de qué probabilidad de tener glaucoma indicaríamos un tratamiento? ¿Cómo hacer este cálculo? Hay dos opciones.

Una es recurrir de nuevo a la calculadora CASPe donde, a partir de la sensibilidad, la especificidad, el riesgo de la prueba diagnóstica, la probabilidad estimada de enfermedad y las utilidades de las distintas opciones que se pueden dar en la clínica real (tratar a un enfermo, tratar a un sano, no tratar a un enfermo y no tratar a un sano) obtenemos este valor. La utilidad es un número entre 0 y 1 (la calculadora CASPe lo solicita entre 0 y 100) con el que medimos el impacto de la situación. Usted decide cómo medirlo: supervivencia, costes, ausencia de complicaciones, unidades arbitrarias, etc. lo que quiera. Nos solicitan el riesgo de la prueba porque también calcula el umbral diagnóstico (probabilidad de enfermedad a partir de la cual se debería solicitar el test) (17).

Una segunda opción, más sencilla, prescinde del umbral diagnóstico y calcula únicamente el umbral terapéutico a partir del daño que puede ocasionar el tratamiento y del beneficio esperado. El UA se establece como un cociente riesgo/ beneficio (fig. 7), por lo que su valor oscila entre 0 y 1. Para calcularlo debemos primero estimar la utilidad o «impacto» de las dos opciones (tratamiento sí y tratamiento no) en una misma escala que toma valores entre 0 y 1. En nuestro escenario, el daño ocasionado por el tratamiento sería el originado por los efectos secundarios del tratamiento con colirios antihipertensivos, cuya frecuencia se ha estimado en un 4-5% (14). En cuanto al beneficio esperado del tratamiento, se considera que evitar el desarrollo de la neuropatía óptica glaucomatosa supone un descenso del 40% en las complicaciones postoperatorias (15). El resultado final para nuestro escenario es que una probabilidad de glaucoma igual o superior al 12,5% sería suficiente para que el oftalmólogo considerara la instauración del tratamiento.

3) Determinar la probabilidad de enfermedad antes de efectuar la prueba

Con los datos del artículo podemos calcular fácilmente la probabilidad pre-prueba: 0,70 (201/284): existen un total de 201 pacientes con la prueba de referencia positiva del total de los 284 pacientes incluidos en el estudio (10).

Umbral de acción = Daño esperado con el tratamiento/mejoría esperada con el tratamiento.
Daño: Frecuencia de efectos adversos (EA) producidos por el tratamiento * impacto de esos EA (en una escala de 0 a 1).
Mejoría= Frecuencia de EA prevenidos por el tratamiento*impacto de esos EA (en la misma escala de 0 a 1)

Cálculo en nuestro escenario
- Efecto adverso del tratamiento.
- Efecto adverso prevenido al evitar el desarrollo de glaucoma.
- Consideramos que el impacto del efecto adverso del tratamiento (valor=1) equivale al impacto de los efectos adversos prevenidos por el tratamiento (valor=1).
- Frecuencia de efectos adversos del tratamiento: 5%(8)
- Frecuencia de efectos adversos prevenidos por el tratamiento: 40%(9)

Umbral de acción = Daño esperado con el tratamiento/Mejoría esperada con el tratamiento
Umbral de acción = 0,05*1/0,4*1= 0,125 (12,5 %)

Fig. 7: Cálculo del umbral de acción.

Recordemos que estas conversiones entre probabilidades pre y post-prueba se realizan automáticamente a partir de la probabilidad preprueba y la LR, mediante el nomograma de Fagan o, más sencillamente en cualquier ordenador personal con la calculadora CASPe que se encuentra accesible en Internet (16,17).

4) Cálculo de la probabilidad posprueba

Consideremos las probabilidades que hemos obtenido en los pasos previos: si la probabilidad de tener glaucoma en un paciente remitido por ese motivo es del 70% (probabilidad pre-prueba) y si el UA para avisar instaurar tratamiento es del 12,5%, en principio se debería instaurar tratamiento siempre, a no ser que dispusiera de una prueba diagnóstica que fuera capaz de modificar esa probabilidad preprueba del 70% hasta una cifra inferior al UA. ¿Es la GDx esa prueba? No lo parece, ya que si la prueba fuera positiva tendría que instaurarse siempre, y si la prueba fuera negativa ... también, puesto que la LR de la GDx no es suficiente para llevar la probabilidad de glaucoma por debajo del 12,5%.

El propio artículo que hemos analizado termina estableciendo en su discusión que aunque la GDx proporciona una buena discriminación entre pacientes con cambios progresivos en la papila y los individuos sanos, su exactitud no es la adecuada para sugerir su utilización como una prueba aislada en el diagnóstico de glaucoma. Y termina con una conclusión que no nos sorprende: La GDx detecta anomalías en pacientes con un diagnóstico previamente confirmado de glaucoma (10).

5) Resolución del escenario clínico

Así pues, la aplicación de este simple test diagnóstico ha modificado nuestra actitud terapéutica en el paciente concreto del escenario.

¿Qué hubiera ocurrido si hubiésemos solicitado la GDx a pesar de un resultado negativo? Pues que incluso con GDx positiva, habría que pensar que se trata de un falso positivo, ya que suponiendo una LR+ de 5 para la GDx la probabilidad de glaucoma pasaría del 2,5% al 11,4%. Ni así habría que instaurar tratamiento (umbral 12,5%). Este es un buen ejemplo del uso inadecuado de la prueba diagnóstica: si un resultado positivo no es capaz de situarnos más allá del umbral de acción, entonces ¿para qué pedir la prueba?

Aunque hemos llegado a una solución de conveniencia, si fuéramos ortodoxos, los pasos a seguir en caso de que la GDx hubiera sido positiva para glaucoma no serían exactamente los mismos. Imaginemos que el paciente hubiera tenido un índice NFI (Nerve Fiber Indicador) de 36 (el índice NFI asigna un valor de 0-100: a mayor valor absoluto mayor probabilidad de glaucoma): en ese caso instauramos tratamiento, pero ¿debemos pedir un campo visual previo, o debemos instaurar tratamiento? Como decíamos antes, cada nueva prueba diagnóstica modifica la probabilidad de enfermedad, y ahora la probabilidad posprueba se ha transformado en la probabilidad preprueba de la perimetría.

El oftalmólogo se enfrenta ahora a un nuevo escenario con un segundo nivel de decisión para el que el planteamiento expuesto vuelve a ofrecer una solución racional. Primero debe determinar su propio UA en el nuevo escenario. La actitud terapéutica que se plantea ahora es si se debe instaurar o no tratamiento antes de solicitar un campo visual. Para instaurar tratamiento, obviamente el umbral de acción debe ser muy superior al 12,5%. Para establecerlo, el oftalmólogo sabe que intervienen otras consideraciones como, por ejemplo, el porcentaje de pacientes no glaucomatosos en los que está dispuesto a asumir tratamiento, o distintas consideraciones económicas.

Por último, el oftalmólogo debe conocer la capacidad de los test diagnósticos de que dispone en el momento del escenario, para cambiar la probabilidad. Cuando llegue ante el enfermo, debe realizar minuciosamente una anamnesis y una exploración clínica. Si, al terminar, se convence de que la probabilidad de glaucoma es del 90% (valor por encima de su UA) debe instaurar tratamiento. Si, por el contrario, la exploración del paciente sólo ha sido capaz de subir la probabilidad de glaucoma al 60%, deberá solicitar una prueba de imagen: aquella que posea el LR+ más alto, para que sea capaz de llevar la probabilidad de glaucoma más allá del umbral.

Moraleja: en la práctica, el rendimiento diagnóstico de un test depende sobre todo de la probabilidad preprueba, de ahí la importancia de aprender a estimar nuestro nivel de «certeza» en cada momento del proceso de diagnóstico. Sólo así utilizaremos las pruebas diagnósticas juiciosamente. Si mi probabilidad preprueba es muy alta o muy baja, la solicitud de una prueba diagnóstica apenas producirá

Tabla V. Variación en la probabilidad posprueba con un test con LR=13

Probabilidad pre-prueba (%)	Probabilidad post-prueba (%)
1	11,6
2	21
10	59,1
25	81,3
50	92,9
75	97,5
90	99,2

cambios, por muy buena que sea la prueba confirmando o descartando la enfermedad (tabla V). Observe cómo los cambios son mínimos con probabilidades preprueba extremas. El rendimiento diagnóstico (cambio en las probabilidades) es máximo en la situación de máxima incertidumbre: el máximo beneficio se produce cuando no somos capaces de inclinarnos ni a favor ni en contra del diagnóstico, es decir, cuando la probabilidad preprueba es del 50%. ¿Qué hacemos cuando preferimos ver a un paciente con papila sospechosa a los 6 meses y pedirle un campo visual? Dejamos pasar un tiempo y volvemos a valorar para ver si su probabilidad preprueba ha aumentado. Quizá entonces nos encontremos por encima del umbral de acción y no sea necesario solicitar nuevas pruebas.

A lo largo de este artículo hemos conocido los requisitos que debe tener un estudio sobre pruebas diagnósticas para que sus resultados sean aceptables. Sin embargo, que una prueba diagnóstica tenga una alta sensibilidad y especificidad no significa que su uso proporcione un rendimiento adecuado en todos los casos. Conocer el umbral de acción, la probabilidad previa de enfermedad en un paciente concreto y la capacidad que tiene el test para modificar esta probabilidad (su LR+ y LR-) serán los requisitos básicos para un uso racional. Recuerde que por bueno que sea el test en cuestión, el rendimiento máximo lo obtendrá cuando se encuentre en situación de máxima incertidumbre o, lo que es lo mismo, estime –en ese momento determinado del proceso diagnóstico– una probabilidad de enfermedad del 50%. Un secreto: use los tests cuando haya aumentado al máximo la probabilidad preprueba, tras una minuciosa anamnesis y exploración física.

BIBLIOGRAFÍA

1. *Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996; 312: 71-72.*
2. *Roberts I. Cochrane injuries group albumin reviewers. BMJ 1998; 317: 235-240.*
3. *Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309: 597-599.*
4. *Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. BMJ 1995; 310: 1085-1086.*
5. *Guyatt G, Cook D, Haynes B. Evidence based medicine has come a long way. BMJ 2004; 329: 990-991.*
6. *Sackett DL. The sins of expertness and a proposal for redemption. BMJ 2000; 320: 1283.*
7. *Weil RJ. The future of surgical research. PLoS Med 2004. Disponible en: www.plosmedicine.org*
8. *Spins and Snouts. Center for evidence based medicine. Disponible en: www.cebm.net/index.aspx?o=1042*
9. *Programa de lectura crítica CASPe: 10 preguntas para entender un artículo sobre diagnóstico. Disponible en: www.redcaspe.org/herramientas*
10. *Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Use of progressive glaucomatous optic disk change as the reference standard for evaluation of diagnostic tests in glaucoma. Am J Ophthalmol 2005; 139: 1010-1018.*
11. *Bland JM, Altmann DG. Measurement error and correlation coefficients. BMJ 1996; 313: 41-42.*
12. *Standards for the reporting of diagnostic accuracy studies. Disponible en: www.stard-statement.org*
13. *Likelihood ratios. Center for evidence based medicine. Disponible en: www.cebm.net/index.aspx?o=1162*
14. *Pre-test probability. Center for evidence based medicine. Disponible en: www.cebm.net/index.aspx?o=1041*
15. *Kirwan JF, Nightingale JA, Bunce C, Wormald R. Beta blockers for glaucoma and excess risk of airways obstruction: population based cohort study. BMJ 2002; 325:1396–1397.*
16. *Glasziou P. Métodos para utilizar el teorema de Bayes en la cabecera del enfermo. Evidence based medicine 2002 Ene-Feb Disponible en: www.evidence-basedmedicine.com*
17. *Calculadora pruebas diagnósticas. Disponible en: www.redcaspe.org/herramientas/descargas/pruebasdiagnosticas.xls*