

MBE en Cirugía Pediátrica. Lectura crítica de artículos. Pruebas diagnósticas (I)

V. Ibáñez Pradas, V. Modesto Alapont¹

Servicio de Cirugía Pediátrica. Hospital General de Castellón.

¹Sección de UCI Pediátrica. Servicio de Anestesiología. Hospital Infantil La Fe, Valencia.

INTRODUCCIÓN. EL PROCESO DIAGNÓSTICO

El diagnóstico es una de las habilidades del médico más apreciadas por los pacientes y por los mismos clínicos, como lo atestigua el reciente éxito de la serie de televisión «House, MD» o el famoso dicho popular: «Iré a este médico; él seguro que me lo acierta». Sin embargo, durante el pregrado el médico recibe escasa formación reglada en este terreno. Si a ello unimos que los artículos sobre diagnóstico suelen adolecer, en general, de una muy pobre calidad metodológica, entenderemos fácilmente un fenómeno creciente: que el clínico suele incorporar acríticamente a sus estrategias diagnósticas las novísimas (y carísimas) pruebas (aunque «prueba diagnóstica» es el término adecuado en castellano, por motivos prácticos utilizaremos también el término anglosajón «test») sin que con ello se aumente significativamente el rendimiento diagnóstico de tales estrategias. Sólo se incrementa desorbitadamente el gasto.

Para aprender a usar de manera eficiente una prueba diagnóstica, debemos entender previamente una serie de requisitos básicos. El primero es que una prueba diagnóstica es útil, desde el punto de vista clínico, sólo si nos induce a tomar las decisiones (terapéuticas) adecuadas. El segundo es que 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. Y el tercero es que sólo tiene sentido plantear el uso de nuevos tests si sabe-

mos que sus resultados van a ser capaces de disminuir nuestra situación de incertidumbre.

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). Ello sólo tiene una solución: practicar, practicar y practicar, etc. El grado de incertidumbre/certeza que tenemos de que ocurra un evento (p. ej., que el paciente que tenemos delante tenga, finalmente, una determinada enfermedad), puede expresarse de dos maneras equivalentes (Fig. 1): 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. 2). Una prueba

Correspondencia: Dr. V. Ibáñez Pradas. Servicio de Cirugía Pediátrica. Hospital General de Castellón. Avda Benicàssim, s.n. 12004 Castellón.

Recibido: Agosto 2006

Aceptado: Septiembre 2006

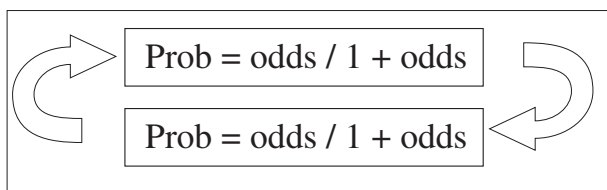


Figura 1. Relación entre odds y probabilidad.

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 «ESpIN», regla mnemotécnica que indica que un resultado positivo (p) diagnóstica (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.

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 cama 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 $p(\text{no}E|-)$, *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 baladí, sino la conocida «falacia de transposición de los condicionales». Expresado de otro modo: la probabilidad de que estés muerto dado que te han ahorcado (muy superior al 90%) no es igual a la

probabilidad de que te hayan ahorcado dado que estás muerto (inferior a una milésima). El valor predictivo global se define como la probabilidad que tiene una prueba de acertar (Fig. 2).

¿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. 3), 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. Así los programas de *screening* de cáncer de mama serán más rentables en EE.UU. que en España, ya que allí la prevalencia de este cáncer es el doble que la nuestra. Otra consecuencia práctica es que los médicos de los niveles terciarios aciertan más con las mismas herramientas que los de primaria, ya que la primaria actúa como un filtro y aumenta la prevalencia de determinados procesos en la asistencia especializada.

Por último, para usar eficientemente una prueba diagnóstica necesitamos el nivel de certeza que queremos tener para llevar a cabo una actitud terapéutica. En el siguiente artículo de la serie intentaremos adiestrar al lector, utilizando para ello el llamado modelo de Pauker-Kassirer.

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. Genéricamente coinciden con las preguntas para artículos sobre tratamiento: ¿son los resultados válidos?, ¿cuáles son?, ¿me serán útiles? En nuestro medio, quizá el artículo sobre prueba diagnóstica más común verse sobre el diagnóstico de la apendicitis, publicado hasta en tres ocasiones en esta revista⁽¹⁻³⁾. Continuaremos con esta patología en el artículo que hemos escogido como ejemplo.

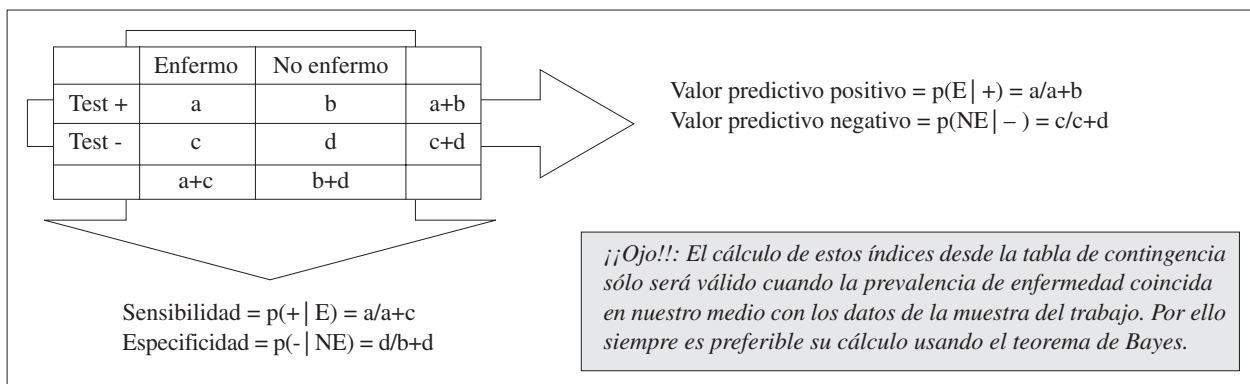


Figura 2. Determinación de la sensibilidad, la especificidad y los valores predictivos a partir de una tabla 2x2.

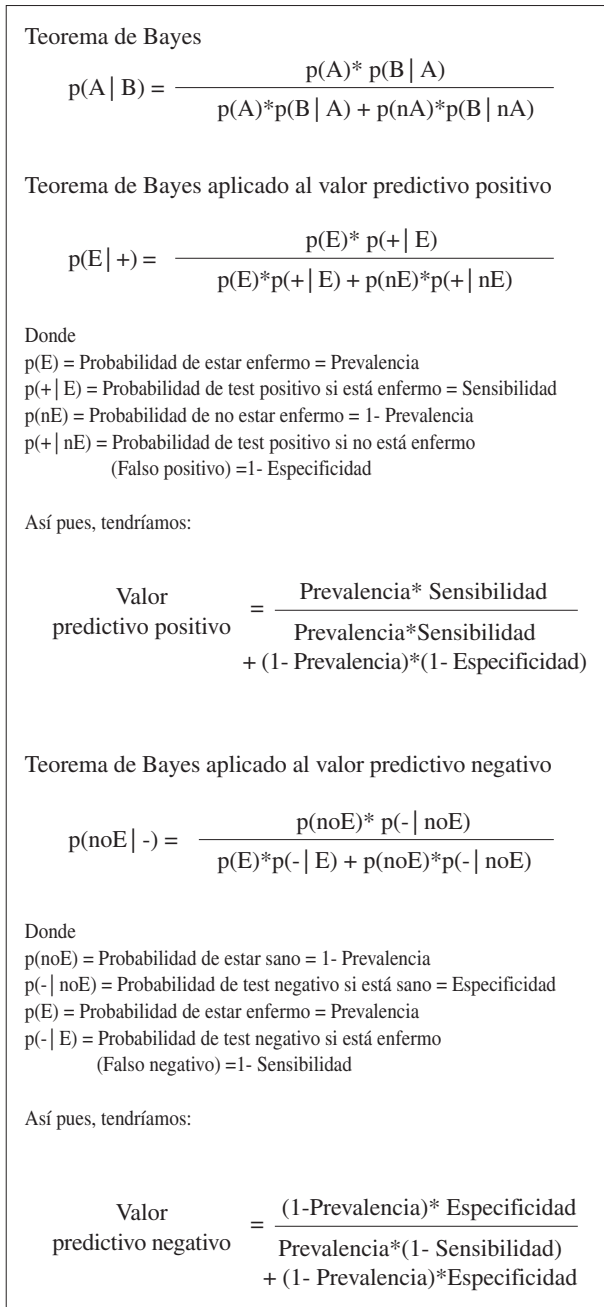


Figura 3. Teorema de Bayes y su aplicación a la determinación del valor predictivo.

Escenario clínico

Usted realiza guardias localizadas como cirujano pediátrico en un hospital de complejidad media cuya puerta de urgencias es atendida por un residente y un adjunto de pediatría. Con el fin de evitar desplazamientos innecesarios, los pediatras han adquirido la costumbre de solicitar ecografías abdominales en los pacientes remitidos por sospecha de apendicitis antes de avisarle, lo que ha motivado quejas por parte del servicio de radiología, que considera que se solicitan las pruebas de imagen de

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?

Tabla II Regla de decisión clínica obtenida mediante regresión logística (Kharbanda y cols.)⁽⁴⁾

Variables	Puntos
Se considera test negativo (no apendicitis) si puntuación ≤ 5	
Neutrófilos absolutos > 6750/μl	6
Dolor a la descompresión o con la percusión	2
Dificultad para caminar	1
Náuseas	2
Migración del dolor a fosa ilíaca derecha	1
Dolor focalizado en fosa ilíaca derecha	2

forma indiscriminada. Una mañana el gerente del hospital le llama a su despacho, donde le espera «un tanto alterado» el jefe del servicio de rayos: han decidido que no van a hacer ecografías que no hayan sido solicitadas expresamente por el cirujano que, previamente, debe haber explorado personalmente al paciente. Con ello pretenden solucionar el problema, pues esperan una petición más juiciosa de la prueba diagnóstica. Aturdido por la alevosía y un tanto soliviantado les comenta que cree tener la solución, y que le dejen una semana antes de cambiar las cosas. Atemorizado por la perspectiva «viajera» que se le avecina realiza una búsqueda bibliográfica y encuentra un trabajo publicado en Pediatrics⁽⁴⁾ en Septiembre de 2005, donde se ofrecen dos reglas de decisión distintas para identificar pacientes con bajo riesgo de tener apendicitis. El artículo obtiene dos modelos realizados mediante metodologías diferentes, uno mediante regresión logística y otro mediante partición iterativa, incluyendo cada uno diferentes variables (Tabla II y Fig. 4).

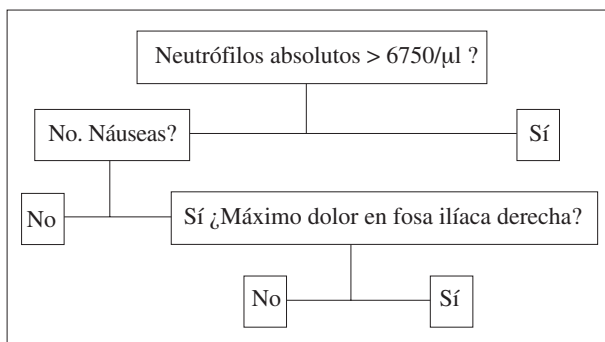


Figura 4. Regla de decisión clínica obtenida mediante partición iterativa (Kharbanda y cols.)⁽⁴⁾.

Ambos modelos se validaron en una muestra de pacientes distintos a los utilizados para generarlos, obteniendo de esta validación los valores diagnósticos. Con los datos del artículo cree que sería posible que los pediatras, antes de llamarle, fueran capaces de identificar certeramente a los pacientes con bajo riesgo de apendicitis, lo que supondría disminuir el número de ecografías solicitadas (y con ello sus viajes vespertinos) sin aumentar el riesgo de peritonitis.

¿SON VÁLIDOS LOS RESULTADOS?

Comparación independiente (y ciega) con una prueba de referencia

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 aquellos 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. Ésta es la opción que utilizaron en nuestro artículo, ya que los pacientes no intervenidos fueron controlados telefónicamente a las 2-4 semanas de su visita a urgencias. También se especifica que para aquellos casos en los que no se pudo localizar a la familia se contactó con el pediatra del paciente para conocer el diagnóstico final.

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 posi-

tivo. Como hemos dicho anteriormente, en los estudios prospectivos este sesgo 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 apendicitis es posible que el radiólogo desconozca datos clínicos del paciente, pero generalmente el cirujano toma su decisión conociendo el resultado de la ecografía, 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 radiólogo 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 trabajo que nos ocupa, los datos se recogieron independientemente de la evaluación del cirujano y antes de que se obtuvieran pruebas de imagen.

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, lo que significa que deberá incluir pacientes con trastornos que puedan simular la enfermedad. 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. Así ocurrió con el antígeno carcinoembrionario (CEA) para la detección del cáncer de colon, estudiado inicialmente con sujetos con cáncer avanzado y sujetos sanos. Al aplicar el test en pacientes con distintos estadios de enfermedad, los resultados empeoraban notablemente, hasta el punto que en la actualidad sólo se utiliza para el seguimiento de pacientes ya diagnosticados⁽⁵⁾. 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 el caso de las apendicitis el rendimiento diagnóstico tanto de la exploración clínica como de la ecografía varía en función de la edad del paciente y de las horas de evolución del cuadro. 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, porcentaje que puede variar en función del tipo de atención primaria de la zona.

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. Por ejemplo, en un artículo sobre la utilidad de la TC para disminuir los ingresos en observación por sospecha de apendicitis, los autores calcularon el rendimiento del test sólo para aquellos casos en los que se visualizó el apéndice, pero eso ocurrió sólo en el 55,7% de la muestra estudiada⁽⁶⁾. Si hubieran incluido estos casos al hacer sus cálculos, el rendimiento del test no hubiera sido tan espectacular, al menos en el subgrupo de pacientes con resultado no concluyente. En estos casos se pueden considerar varias categorías de resultado y calcular el rendimiento diagnóstico para cada categoría. Así, en un estudio sobre ecografía y apendicitis realizado en el hospital de Cruces⁽⁷⁾, los autores distinguían tres diagnósticos ecográficos: de certeza, probable y poco probable, como ocurre en la práctica real. Sin embargo, cuando calcularon el rendimiento del test, no se distinguió entre estas categorías, perdiendo así información al transformar un test con varias opciones en otro dicotómico (sano/enfermo).

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? Fueron pacientes remitidos por dolor abdominal y sospecha de apendicitis. En su centro de trabajo, antes de solicitar una prueba de imagen el paciente ha de ser valorado por un cirujano. Los pacientes estaban comprendidos entre los 3 y los 18 años, y de los 654 pacientes elegibles, se incluyeron 601 (92%). La prevalencia de apendicitis en la muestra fue del 35%, con un 18% de perforaciones; cifras similares a las de nuestro medio^(2,3) (en el estudio cifran las perforaciones en un 22%, pero este porcentaje no coincide con las cifras que se ofrecen: 38 perforaciones en 211 casos intervenidos). Se completó el seguimiento en 593 pacientes (98,6%). En principio y con estos datos, la población se asemeja bastante a la que puede encontrar nuestro colega «imaginario».

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. Es posible que usted lea un artículo sobre los buenos resultados de la TC en el diagnóstico de la apendicitis, pero si los buenos resultados dependen de la utilización de contraste oral o rectal⁽⁸⁾ en la mayor parte de los casos quizá la aplicabilidad de la prueba en su medio de trabajo en concreto se vea disminuida.

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⁽⁹⁾. En el artículo que estamos revisando no se ofrece ninguna información de este tipo y, aunque la valoración de los neutrófilos absolutos no plantearía problema alguno, tal vez la valoración clínica de la fosa ilíaca derecha conlleve cierto grado de subjetividad.

bilidad 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⁽⁹⁾. En el artículo que estamos revisando no se ofrece ninguna información de este tipo y, aunque la valoración de los neutrófilos absolutos no plantearía problema alguno, tal vez la valoración clínica de la fosa ilíaca derecha conlleve cierto grado de subjetividad.

¿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.

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 encontrarse 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⁽¹⁰⁾. Sin embargo, esta definición es la que se ha popularizado al ser la elegida por el grupo español de medicina basada en la evidencia. 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. 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. 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).

Todo eso está muy bien, pero ¿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 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

Razón de verosimilitud positiva (RVP) o <i>likelihood ratio</i> (LR) para test positivo =	Sensibilidad 1 - Especificidad
Razón de verosimilitud negativa (RVN) o <i>likelihood ratio</i> (LR) para test negativo =	1 - Sensibilidad Especificidad
Odds pre-test * LR = Odds post-test	

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

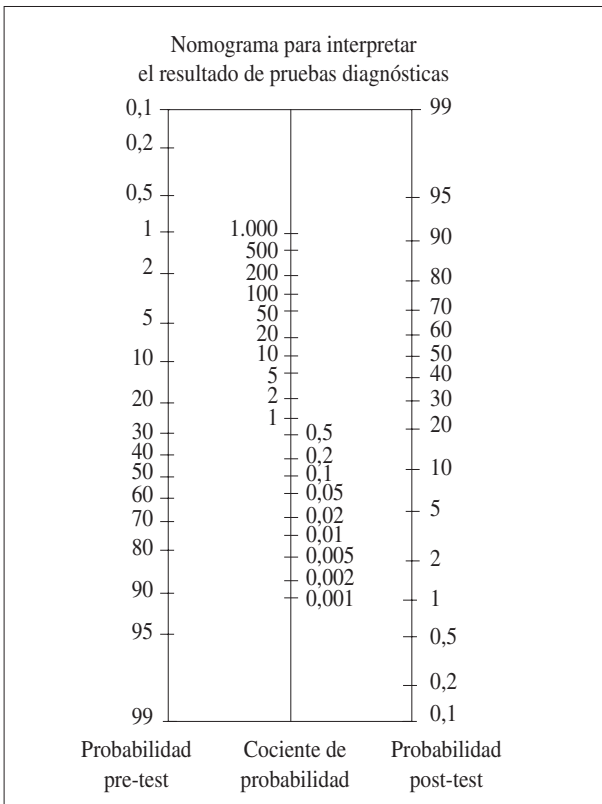


Figura 6. Nomograma de Fagan (adaptado). Nótese que se utiliza la expresión «cociente de probabilidad» para referirse al cociente de verosimilitud o *likelihood ratio*.

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 una apendicitis? De forma genérica, el riesgo a lo largo de nuestra vida estaría en torno al 7-8,6%, que equivaldría a su incidencia^(11,12); entre los pacientes que son derivados a la puerta de urgencias de su hospital y que Vd. tiene que valorar por dolor abdominal estaría en torno al 30%⁽²⁻⁴⁾ pero si usted atien-

Tabla III Interpretación clínica de la razón de verosimilitud (*likelihood ratio*): capacidad para modificar la probabilidad preprueba

LR+ > 10	Aumento grande: Test excelente
LR+ entre 5 y 10	Aumento moderado: Test bueno
LR+ entre 2 y 5	Aumento pequeño: Test malo
LR+ < 2	Aumento insignificante: Test inútil
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

de a un escolar de 6 años con dolor de 24 horas de evolución, inicialmente localizado en mesogastrio y con posterior migración a fosa ilíaca derecha (FID), con defensa a la palpación selectiva en FID y descompresión dolorosa, su estimación de la probabilidad de apendicitis posiblemente sea superior al 80%. 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⁽¹³⁾ (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 posttest. 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 posttest.

Cuanto más extremos sean los valores 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 III 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. En el próximo capítulo calcularemos estos valores a partir de los datos proporcionados por los autores del estudio que estamos analizando.

BIBLIOGRAFÍA

1. Valladares JC, Alaminos M, Castejón J, Fernández R, Sánchez C, Gassó C, et al. Utilidad de la ultrasonografía para el diagnóstico de apendicitis aguda en la infancia. *Cir Pediatr* 1998;**11**:67-70.
2. Valladares JC, Rabaza MJ, Martínez S, Alaminos M, Jiménez C. Eficacia de la ecografía en el diagnóstico de la apendicitis aguda. *Cir Pediatr* 2000;**13**:121-123.
3. Gallinas F, Garde C, Pérez A, Bento L, Martínez MA, Conde J, et al. La ecografía en la selección del dolor abdominal quirúrgico urgente. Estudio prospectivo. *Cir Pediatr* 2004;**17**:141-144.
4. Kharbanda AB, Taylor GA, Fishman SJ, Bachur RG. Decision rule can help identify children at lower risk for appendicitis. *Pediatrics* 2005;**116**:709-716.
5. Bates SF. Clinical applications of serum tumor markers. *Ann Intern Med* 1991;**115**:623-638.
6. Acosta R, Crain FF, Goldman HS. CT can reduce hospitalization for observation in children with suspected appendicitis. *Pediatr Radiol* 2005;**35**:495-500.
7. Vázquez MA, Morteruel E, García E, Mintegui S, Canapé S, Benito J. Rendimiento de la ecografía abdominal en el diagnóstico de apendicitis aguda. *An Pediatr (Barc)* 2003;**58**:556-561.
8. Sivit CJ, Dudgeon DL, Applegate KE, Borisa VJ, Berlin SC, Morrison SC, et al. Evaluation of Suspected Appendicitis in Children and Young Adults: Helical C. *Radiology* 2000;**216**:430-433.
9. Bland JM, Altman DG. Statistical method for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307-310.
10. Ibáñez V, Modesto V, Tosca R. Sobre las likelihood ratios. *Med Clin (Barc)* 2001;**116**:396.
11. Rothrock SG, Pagane J. Acute appendicitis in children: emergency department diagnosis and management. *Ann Emerg Med* 2000;**36**:39-51.
12. Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol* 1990;**132**:910-925.
13. Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med* 1975;**293**: 257.