



Fundamentos de medicina basada en la evidencia

Evaluación de artículos científicos sobre pruebas diagnósticas

Carlos Ochoa Sangrador. Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. (España). Correo electrónico: cochoas@meditex.es

Javier González de Dios. Departamento de Pediatría. Hospital de Torrevieja. Universidad Miguel Hernández. Alicante (España). Correo electrónico: jgonzalez@torrevieja-salud.com

José Cristóbal Buñuel Álvarez. ABS Girona-4. Institut Català de Salut. Girona (España). Correo electrónico: p416ucua@pgirona.scs.es

Términos clave en inglés: evidence-based medicine; diagnosis

Términos clave en español: medicina basada en pruebas; diagnóstico

Fecha de recepción: 10 de febrero de 2007

Fecha de aceptación: 15 de febrero de 2007

Fecha de publicación: 1 de marzo de 2007

Evid Pediatr. 2007; 3: 24 doi: vol3/2007_numero_1/2007_vol3_numero1.24.htm

Cómo citar este artículo

Ochoa Sangrador C, González de Dios J, Buñuel Álvarez JC. Evaluación de artículos científicos sobre pruebas diagnósticas. Evid Pediatr. 2007; 3:24.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín por medio del ETOC <http://www.aepap.org/EvidPediatr/etoc.htm>

Este artículo está disponible en: http://www.aepap.org/EvidPediatr/numeros/vol3/2007_numero_1/2007_vol3_numero1.24.htm
EVIDENCIAS EN PEDIATRIA es la revista oficial del Grupo de Pediatría Basada en la Evidencia de la Asociación Española de Pediatría de Atención Primaria. © 2005-06. Todos los derechos reservados

Evaluación de artículos científicos sobre pruebas diagnósticas

Carlos Ochoa Sangrador. Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. (España).

Correo electrónico: cochoas@meditex.es

Javier González de Dios. Departamento de Pediatría. Hospital de Torrevieja. Universidad Miguel Hernández. Alicante (España). Correo electrónico: jgonzalez@torrevieja-salud.com

José Cristóbal Buñuel Álvarez. ABS Girona-4. Institut Català de Salut. Girona (España).

Correo electrónico: p416ucua@pgirona.scs.es

Aproximación al concepto de diagnóstico médico en la toma de decisiones clínicas

El diagnóstico médico es un proceso dinámico en el que se intenta tomar decisiones idóneas en presencia de incertidumbre. Desde un punto de vista funcional, consideramos prueba diagnóstica a cualquier procedimiento realizado para confirmar o descartar un diagnóstico o incrementar o disminuir su verosimilitud.

El área de investigación en pruebas diagnósticas es, junto con el de innovaciones terapéuticas, una de las áreas con mayor y más rápido desarrollo en la medicina actual. Es posible que este rápido desarrollo de nuevos procedimientos diagnósticos no haya tenido un paralelismo en el desarrollo de las más adecuadas estrategias de evaluación.

Los rápidos avances en los métodos diagnósticos implican la necesidad de que los profesionales dispongan de la información correcta sobre sus características y la aplicabilidad en su ámbito de trabajo. Así pues, la investigación sobre pruebas diagnósticas tiene como objetivos estimar la capacidad discriminatoria de una prueba diagnóstica entre enfermos y no enfermos (sensibilidad-especificidad), determinar el rendimiento de la misma (valores predictivos) o evaluar la utilidad y satisfacción de un procedimiento diagnóstico. La utilidad de una prueba diagnóstica depende de su capacidad de producir los mismos resultados cada vez que se aplica en similares condiciones (fiabilidad) y de que sus mediciones reflejen exactamente el fenómeno que se intenta medir (validez o exactitud), pero también de su rendimiento clínico y de su coste¹⁻³.

Si evaluamos nuestra práctica clínica observaremos que a menudo empleamos procedimientos diagnósticos de los que desconocemos estos parámetros. Para conocer esta información debemos recurrir a los estudios de evaluación de pruebas diagnósticas publicados. Sin embargo, en muchos de estos estudios no siempre encontramos información válida, relevante o aplicable a nuestro entorno clínico⁴⁻⁷.

En esencia, realizar el diagnóstico es asignar con razonable incertidumbre (es decir, con razonable probabilidad) un paciente a una clase (o grupo) constituida por sujetos con una enfermedad o entidad nosológica. Por lo tanto, el diagnóstico es el primer paso, sin duda crucial, que nos permite la utilización de otras evidencias en la toma de decisiones sobre el paciente. El diagnóstico no es un fin en sí mismo, sino un instrumento en la toma de decisiones clínicas; de hecho no es preciso tener una seguridad diagnóstica absoluta para adoptar la decisión terapéutica correcta^{8,9}.

El modelo de razonamiento que subyace en un estudio sobre pruebas diagnósticas podría esquematizarse del siguiente modo: existe un fenómeno clínico que puede ser medido de forma fiable y válida por un procedimiento que se llamará diagnóstico o gold standard. En este marco nos planteamos dos tipos de pregunta: ¿hay un segundo procedimiento de medida, que llamaremos test o prueba, que podría medir también este fenómeno, de modo fiable y válido? y ¿ese segundo procedimiento tiene algún tipo de ventaja respecto al primero? Las ventajas del segundo método podrán ser teóricas, es decir, que mejore la validez y precisión del primer procedimiento (en cuyo caso estamos buscando un nuevo gold standard) o prácticas, es decir, que sea más fácil o económico, con menos riesgo o molestias, etc (en cuyo caso estamos buscando un procedimiento que evite realizar el gold standard). Cualquier medición tiene dos componentes básicos, un concepto a medir y un procedimiento del que cabe destacar tres elementos: el protocolo para proceder a la medición, el resultado expresable en alguna escala y los criterios de interpretación.

Existe un planteamiento erróneo al realizar estudios de evaluación de pruebas diagnósticas: al aplicar una prueba diagnóstica en mi paciente, si es (+) estará enfermo y si es (-) estará sano. El planteamiento real sería el siguiente: mi paciente tiene ya una probabilidad de estar enfermo (probabilidad preprueba, que influye mucho en el resultado); al aplicar una prueba diagnóstica, si es (+) está probabilidad de enfermedad será mayor y si es (-) será menor (probabilidad postprueba). Para el cálculo de estas probabilidades emplearemos los cocientes de probabilidades (CP), que en el escenario más simple serán los CP positivo (mayor de 1) y negativo (menor de 1). En la figura 1 exponemos los aspectos fundamentales de los estudios sobre evaluación de pruebas diagnósticas.

Existen determinados aspectos en el diseño de un estudio sobre pruebas diagnósticas que pueden afectar a la precisión (errores aleatorios) o a la validez (errores sistemáticos) de las estimaciones realizadas, y de este modo pueden amenazar la calidad de la evidencia que aporta el estudio¹⁰. En la tabla 1 se resumen los sesgos potenciales en los estudios sobre pruebas diagnósticas⁸.

Si queremos tomar las mejores decisiones en la elección y aplicación de pruebas diagnósticas para nuestros pacientes, tenemos que estar preparados para integrar nuestros conocimientos y experiencia previos con la información científica de los trabajos que continuamente vienen incorporándose a la literatura médica. Para ello, necesitamos estar familiarizados con los parámetros que

describen la validez y fiabilidad de las pruebas diagnósticas y aprender a evaluar la validez, relevancia y aplicabilidad de los estudios donde son estimados^{11,12}.

Evaluación de una prueba diagnóstica

Siguiendo las recomendaciones del Evidence-Based Medicine Working Group^{13,14} analizaremos los tres pasos a seguir para analizar la VALIDEZ, la IMPORTANCIA y la APLICABILIDAD (Tabla 2) de los artículos sobre pruebas diagnósticas, planteando diversas cuestiones^{1,15,16}.

¿Son válidos los resultados?

¿Cuáles son los resultados?

¿Son aplicables en tu medio?

1.- ¿Son válidos los resultados del estudio?

Cabe contestar a una serie de preguntas, clasificadas en criterios primarios (son preguntas de eliminación: si no se cumplen estos criterios primarios quizá no valga la pena continuar con la lectura) y secundarios (son preguntas más detalladas, que cabe realizar si se continúa con la lectura del artículo tras cumplir con los criterios primarios).

a.- Criterios primarios:

¿Existió una comparación ciega, independiente, con un patrón de referencia adecuado?: para saber si una prueba diagnóstica es útil sus resultados deberían compararse con la "verdad" (presencia o ausencia de enfermedad), en la práctica con otra prueba que actúa como patrón de referencia (gold standard), el cual debe estar reconocida como tal por la comunidad científica (por contar con una validez contrastada o, al menos, aceptada por consenso); este patrón de referencia se debe aplicar a toda la serie de casos estudiados y no tiene que incorporar información procedente de la prueba diagnóstica que se evalúa (sesgo de incorporación: evitar que el resultado de una de las pruebas pueda influir sobre la interpretación de los resultados de la otra). Si no se puede aceptar el estándar de referencia dentro de lo razonable, es poco probable que el artículo proporcione resultados válidos para sus objetivos.

En relación con el patrón de referencia, resulta también importante considerar si es capaz de clasificar el estado de enfermedad en todas las observaciones. En el caso de que existan observaciones con un diagnóstico indeterminado, si éstas son excluidas del análisis, se producirán estimaciones sesgadas de las características operativas de la prueba diagnóstica. Este sesgo, conocido como sesgo por exclusión de indeterminados, ocasiona habitualmente sobrestimaciones de la sensibilidad y de la especificidad.

¿Incluyó la muestra de pacientes a un espectro adecuado de pacientes a los cuales en la práctica clínica se aplicará la prueba diagnóstica?: el valor real de una prueba diagnóstica sólo se establece en un estudio que es lo más parecido a la práctica clínica, por lo que es conveniente que incluya pacientes con sospecha de enfermedad, que incluyan sujetos enfermos (con distinta gravedad) y sanos. El artículo debe definir explícitamente los criterios de selección ya que, si bien la sensibilidad y especificidad

son relativamente constantes independientemente de la prevalencia de la enfermedad, los valores predictivos, y por tanto la utilidad de la prueba para el lector, están muy influidos por ella.

b- Criterios secundarios:

¿Influyeron los resultados de la prueba diagnóstica en la decisión de realizar el patrón de referencia?: conviene evitar los sesgos de secuencia y de revisión.

Evitar el sesgo de secuencia (verificación diagnóstica o verification bias): el diseño del estudio debe tratar de garantizar que en la muestra no se hayan excluido pacientes, en función del resultado de la prueba o de la existencia de mayor o menor riesgo de enfermedad. Incurriríamos en un sesgo de verificación diagnóstica cuando la probabilidad de que se les realice el patrón de referencia sea menor entre los sujetos con la prueba diagnóstica negativa y por lo tanto sea menos probable que éstos entren en el estudio. Todos los sujetos deben haber sido sometidos tanto a la nueva prueba como a la de referencia; sin embargo, en algunos estudios (sobre todo cuando el patrón de referencia es una prueba invasiva) solamente se aplica a los sujetos en los que se ha obtenido un resultado positivo de la prueba que se estudia. En los estudios de cohortes se acepta el criterio si a todos los sujetos se les practica la prueba diagnóstica y el patrón de referencia; en los estudios de casos y controles en los que la prueba diagnóstica precede a la prueba de referencia, se cumple el estándar si la verificación de la enfermedad se realiza en una serie consecutiva de pacientes con independencia del resultado de la prueba diagnóstica; en los estudios de casos y controles en los que la prueba de referencia precede a la prueba diagnóstica, el crédito se obtiene cuando los resultados de ésta son estratificados de acuerdo con los factores clínicos que potencialmente condicionan la indicación de la prueba de referencia.

Evitar el sesgo de revisión (valoración ciega o review bias): tanto la nueva prueba como la de referencia deben haber sido realizadas de forma independiente y ciega, para evitar la interpretación subjetiva de una prueba influida por el conocimiento del diagnóstico o de las características clínicas del paciente. Para estudios de cohortes prospectivos en los que el paciente recibe primeramente la prueba diagnóstica, el estándar es aceptado si la prueba de referencia se evalúa independientemente, es decir, sin conocer el resultado de la prueba diagnóstica; en estudios prospectivos donde la prueba de referencia precede a la prueba diagnóstica y en los estudios de casos y controles, se cumple el estándar si figura una afirmación de la independencia de interpretación de los datos.

¿Se describieron los métodos para realizar el examen con el suficiente detalle para permitir su reproducción?: valorar si es adecuada la descripción del procedimiento de realización de la prueba diagnóstica, el estudio de su reproducibilidad y la definición del término normal y/o anormal, así como la justificación del criterio de normalidad elegido, y el análisis e interpretación de sus resultados.

Si después de considerar todos estos aspectos hemos

decidido que el estudio es suficientemente válido, procederemos a examinar las propiedades de la prueba diagnóstica.

2- ¿Cuáles son los resultados del estudio?

El proceso diagnóstico es en esencia un cálculo de probabilidades. La utilidad de una prueba diagnóstica depende de su validez y de su fiabilidad. La finalidad del análisis es cuantificar la capacidad de una prueba diagnóstica para clasificar correcta o incorrectamente a una persona según la presencia o ausencia de una enfermedad.

El punto de partida del proceso diagnóstico es habitualmente un paciente, con unas características de gravedad y comorbilidad concretas, que le confieren una probabilidad determinada de tener la entidad a diagnosticar (probabilidad preprueba). El objetivo de la realización de la prueba diagnóstica es, una vez conocido el resultado, modificar esa probabilidad hasta obtener una probabilidad postprueba. La magnitud y dirección de ese cambio va a depender de las características operativas de la prueba diagnóstica, pero en todo caso debemos tener en cuenta que el punto de partida, la probabilidad preprueba, va a resultar muy importante en ese proceso.

¿Se han presentado los cocientes de probabilidad o los datos para calcularlos?

Para poder evaluar la validez de una prueba diagnóstica se requiere un patrón de referencia; el grado de acuerdo entre la prueba diagnóstica y el patrón de referencia puede ser representado en una tabla de contingencia, de donde se obtiene la sensibilidad (S), especificidad (E), valor predictivo positivo (VPP) y valor predictivo negativo (VPN).

La S considera la validez de la prueba entre los enfermos (corresponde a la probabilidad de que un individuo con la enfermedad tenga un resultado positivo de la prueba) y la E la validez de la prueba entre los sanos (corresponde a la probabilidad de que un individuo sin la enfermedad tenga un resultado negativo de la prueba). Los valores complementarios de la S y la E corresponden a las probabilidades de los resultados falsos negativos y falsos positivos, respectivamente. Cuando una prueba tiene una S muy alta, un resultado negativo sirve para descartar la enfermedad (la probabilidad de un falso negativo es muy pequeña); si la prueba tiene una E muy alta, un resultado positivo es prácticamente diagnóstico (la probabilidad de un falso positivo es muy pequeña).

S y E son propiedad intrínseca de las pruebas diagnósticas, que presentan una relación recíproca entre ambas y cuyos resultados son independientes de la prevalencia de la enfermedad, a diferencia de los valores predictivos: así, cuanto mayor sea la prevalencia de la enfermedad mayor será el VPP y menor el VPN; los valores predictivos tienen una utilidad postprueba.

Todos estos estimadores de validez pueden ser aplicados a pruebas con resultados discretos con más de dos categorías e incluso a resultados expresados en variables continuas. En este caso podemos establecer puntos de

corte para convertir el resultado en una variable discreta con dos o más categorías. Otra alternativa que permite explorar la capacidad diagnóstica de una prueba en sus distintos valores son las curvas ROC (iniciales del término inglés original Receiver Operating Characteristics), con las que podemos conocer su validez global y seleccionar el punto o puntos de corte más adecuados. Es frecuente encontrar trabajos en los que se selecciona un punto de corte desplazado de su punto medio, con la intención de maximizar sensibilidad a expensas de reducir especificidad, o viceversa; en estos casos, podemos estar limitando la aplicabilidad de la prueba ya que dichos valores sólo le serán útiles a una parte de los pacientes. Las curvas ROC permiten obtener un gráfico en el que se observan todos los pares S/E resultantes de la variación continua de los puntos de corte; las curvas ROC permiten explorar la capacidad diagnóstica de la prueba en sus distintos valores, de manera que podamos conocer su validez global y seleccionar el punto o puntos de corte más adecuados.

Una forma alternativa de describir el comportamiento de una prueba diagnóstica son los cocientes de probabilidades (CP) -o razón de verosimilitudes, en inglés, likelihood ratio-, definidos como la probabilidad de dicho resultado en presencia de enfermedad divididas por la probabilidad de dicho resultado en ausencia de enfermedad. Para el escenario más simple tenemos dos tipos: cociente de probabilidades de un resultado positivo (CP+) = $S/(1-E)$ y cociente de probabilidades de un resultado negativo (CP-) = $(1-S)/E$. Dado que los CP relacionan la S y la E, no varían con la prevalencia. Los CP adoptan valores entre 0 e infinito, siendo el valor nulo el 1, y cuya interpretación se valora según los criterios expresados en la tabla 3.

Consideremos el escenario diagnóstico más simple, en el que tanto el patrón de referencia como la prueba diagnóstica clasifican a los pacientes en dos grupos, en función de la presencia o ausencia de un síntoma, signo o enfermedad. Utilizando los CP se pueden calcular las probabilidades postprueba (valores predictivos) a partir de la probabilidad preprueba de cada paciente individual (Tabla 1), que habitualmente no es la misma que la existente en los estudios publicados. La relevancia de la prueba va a depender de cuánto sea capaz de incrementar o disminuir esa probabilidad, una vez aplicados los CP positivo o negativo. Generalmente CP positivos cercanos a 10 y negativos cercanos a 0,1 resultan clínicamente relevantes, aunque finalmente serán las probabilidades postpruebas ajustadas a mi paciente las que indicarán el grado de relevancia diagnóstica.

El CP es una medida de gran utilidad en la práctica clínica, ya que la principal utilidad es que permite calcular la probabilidad postprueba (P_{post} = probabilidad de que un sujeto tenga la enfermedad si obtiene un resultado positivo en la prueba o VPP ajustado) a partir de cualquier probabilidad preprueba (P_{pre} = probabilidad de que un sujeto tenga la enfermedad antes de aplicarle la prueba) o prevalencia, y de esta forma evaluar cuánto puede ganarse si se realiza. El objetivo de la realización de la prueba

diagnóstica es, una vez conocido el resultado, modificar esta Ppre hasta obtener una Ppost; la magnitud y dirección de ese cambio va a depender de las características operativas de la prueba diagnóstica, pero en todo caso debemos tener en cuenta que el punto de partida, la Ppre, va a resultar muy importante en ese proceso.

Para poder operar con los CP en el cálculo de probabilidades, éstas deben transformarse en ventajas (odds); los pasos a seguir son: 1) transformar la Ppre en odds preprueba = $Ppre/1-Ppre$; 2) obtener la odds postprueba = $CP+ \times$ odds preprueba; 3) obtener la Ppost = $odds \text{ postprueba} / 1 + odds \text{ postprueba}$. La diferencia que exista entre la Ppre y Ppost informa de la utilidad que tiene una determinada prueba diagnóstica; a mayor diferencia entre una y otra probabilidad mayor contribución de la prueba al proceso diagnóstico. Afortunadamente existen un modo más simple de realizar estos cálculos (nomograma de Fagan-ver figura 2).

Una de las ventajas de los CP es que si la prueba tiene más de dos resultados posibles, se puede calcular un CP para cada uno de ellos, permitiéndonos interpretar la contribución al diagnóstico de cada resultado. Otra de las ventajas radica en que los CP facilitan el cálculo de las modificaciones de probabilidad obtenidas al aplicar en serie varias pruebas diagnósticas, recurso frecuentemente empleado en la práctica clínica y en los estudios de análisis de decisión.

En la tabla 4 se expresan los indicadores de validez de una prueba diagnóstica, a través de los resultados de un estudio hipotético sobre la utilidad de una prueba diagnóstica, utilizando su patrón de referencia (tabla 4).

En este proceso de cálculo, puede resultar problemático estimar adecuadamente la Ppre o prevalencia; debería hacerse a partir de la propia experiencia acumulada, de forma que sea específica del entorno de trabajo; sin embargo, habitualmente se dispone de poca información sobre estos aspectos; cuando no se está muy seguro de la fiabilidad de la estimación, puede ser conveniente analizar las implicaciones sobre la práctica que puede tener el considerar diferentes valores plausibles de las Ppre.

¿Cuán fiables son los resultados?: la fiabilidad de una prueba viene determinada por la estabilidad de sus mediciones cuando se repite en condiciones similares. La variabilidad de las mediciones va a estar influida por múltiples factores que interesa conocer y controlar. Entre ellos, tiene especial importancia distinguir las variaciones de interpretación intraobservador e interobservador. La fiabilidad puede ser evaluada para resultados discretos nominales mediante el índice kappa (IK), para resultados discretos ordinales mediante el índice kappa ponderado y para resultados continuos mediante el coeficiente de correlación intraclase y el método de Bland-Altman. El IK nos ofrece una estimación del grado de acuerdo no debido al azar a partir de la proporción de acuerdo observado (Po) y la proporción de acuerdo esperado (Pe) = $Po - Pe / 1 - Pe$.

El IK puede adoptar valores entre -1 y +1: es 1 si existe un acuerdo total, 0 si Po es igual a Pe y menor de 0 si Po es inferior a Pe por azar. La interpretación más aceptada de

los rangos de valores del IK son: 0,81-1 = excelente, 0,61 – 0,80 = buena, 0,41 – 0,60 = moderada, 0,21 – 0,40 = ligera, $\leq 0,20$ = mala.

¿Cuán precisos son los resultados?: al igual que en otros tipos de estudios, la valoración de la validez de las pruebas diagnósticas se hace sobre muestras, por lo que los resultados obtenidos son sólo estimaciones puntuales, sujetas a variabilidad aleatoria, y por lo tanto deben expresarse con sus intervalos de confianza (IC); estos IC tendrán que ser aplicados en el cálculo de la Ppost para poder juzgar la utilidad de la prueba diagnóstica.

Los estudios de evaluación de pruebas diagnósticas proporcionan conocimiento imprescindible para el uso de la probabilidad en el diagnóstico. Para que este conocimiento sea válido e importante se requiere saber los posibles errores a evitar y un diseño específico que optimice el esfuerzo investigador, y para que se incorpore este conocimiento a la práctica clínica es preciso disponer de la capacidad de juicio que permita una lectura crítica de las publicaciones sobre pruebas diagnósticas, tal como se ha pretendido en este capítulo.

3- ¿Son aplicables en tu medio?

El hecho de que una prueba diagnóstica determinada sea apropiada, no significa que todos los pacientes puedan utilizar el diagnóstico. Para ellos nos podemos preguntar¹⁷:

¿La prueba está disponible y es reproducible, tanto en la realización como en la interpretación, para los pacientes de mi ámbito?: las propiedades de la prueba pueden cambiar según se apliquen a grupos de sujetos con una distribución diferente de estadios de la enfermedad o de otras enfermedades concomitantes que puedan interferir en el diagnóstico. Cuando se aplica a pacientes con la enfermedad en estadios avanzados el CP+ se aleja de 1 (aumenta la sensibilidad), mientras que si son de estadios menos avanzados, el CP+ se aproxima a 1 (disminuye la sensibilidad). Si los pacientes de mi ámbito cumplen los criterios de inclusión y exclusión de los sujetos del estudio, los resultados de la prueba serán aplicables; pero si no los cumplen, deben valorarse los motivos por los que podría pensarse que los resultados no son aplicables; para la generalización de los resultados siempre es útil disponer de alguna revisión que interprete la información de diferentes estudios.

Obviamente, no puede aplicarse en la práctica una prueba que no esté disponible ni sea asequible en el propio entorno. Pero aunque lo sea, es conveniente evaluar su reproductibilidad, de la que depende en gran medida su utilidad. Debe constar la reproductibilidad de la prueba en el artículo, especialmente cuando ésta comporte cierto grado de subjetividad en su realización o interpretación.

¿Puede generarse una estimación clínicamente razonable de la Ppre de nuestro paciente?: ya ha sido comentado que la utilidad del CP depende de que pueda disponerse de una estimación adecuada de la Ppre de un paciente concreto. La situación ideal es aquella en la que se dispone de datos del propio centro, pero en otras ocasiones esta

estimación puede obtenerse de la propia experiencia o de la especulación clínica, o valorando las diferencias entre las características de nuestro paciente y las de los sujetos estudiados (en este último caso es conveniente realizar análisis de sensibilidad, aplicando varias Ppre clínicamente razonables y determinando su impacto sobre la utilidad de la prueba).

¿La Ppost resultante puede hacer cambiar la actuación en la práctica clínica y ayudar a nuestro paciente?: si la prueba fuera negativa o se obtuviera un CP negativo sensiblemente inferior a 1, la Ppost podría ser tan reducida que hiciera descartar el diagnóstico y no realizar más pruebas; si la prueba fuera positiva u ofreciera un CP+ muy elevado, la Ppost podría ser tan elevada que se consideraría diagnóstica y tampoco sería necesario realizar más pruebas; si la prueba condujera a una situación intermedia entre las dos anteriores, debería plantearse la necesidad de realizar nuevas pruebas. Este cálculo es adecuado siempre que las pruebas sean independientes, hecho que suele ser difícil de asumir en la práctica, por lo que el lector deberá ser más exigente a la hora de interpretar los valores de la Ppost.

La utilidad clínica de la prueba depende también de la repercusión que tengan sobre nuestra actitud diagnóstica y terapéutica; en ocasiones la Ppre de nuestro paciente será tan baja que, independientemente cuál sea el resultado de la prueba, la Ppost será igualmente baja y por lo tanto no merecerá la pena llevarla a cabo (umbral diagnóstico); en el otro extremo, si la Ppre es muy elevada, su resultado no va a cambiar nuestra decisión de tratar (umbral terapéutico), por lo que a veces podremos obviarla; es en la zona intermedia donde más interés debería tener la prueba, pues el grado de información que nos aporta dependerá de la magnitud de los CP: valores cercanos a 1 resultarán poco útiles, mientras que valores lejanos modificarán de forma importante la Ppost y su rendimiento diagnóstico.

Pero el criterio último de la utilidad de una prueba, al margen de que ésta ofrezca información diagnóstica no disponible previamente o de que modifique nuestro comportamiento clínico, es si el paciente obtiene algún beneficio. Existen escenarios en los que la prueba diagnóstica no resulta coste-efectiva, conlleva riesgos, o conduce a decisiones terapéuticas sin repercusión sobre el paciente.

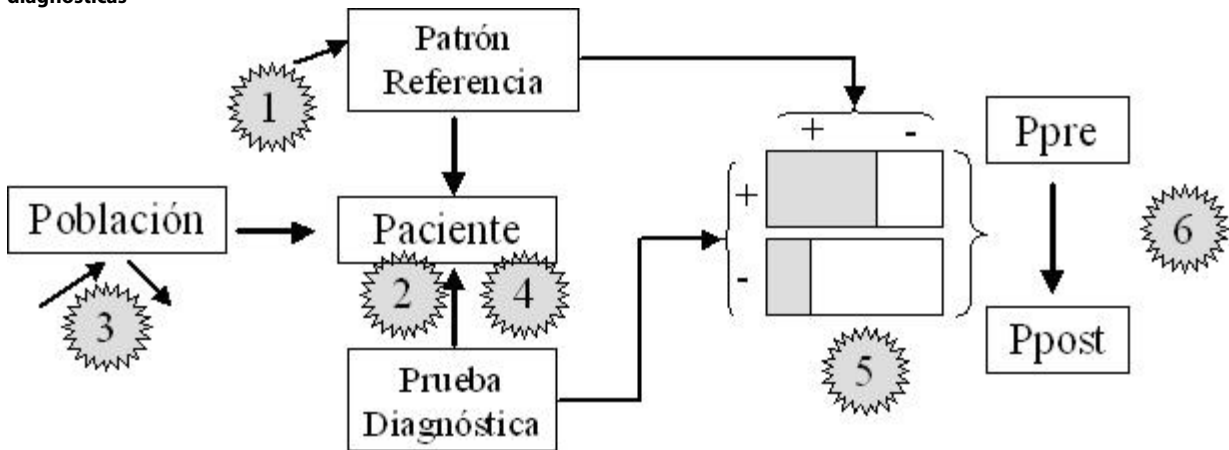
A la hora de decidir qué prueba diagnóstica se va a utilizar, es conveniente tener en cuenta los valores y preferencias del paciente, ya que algunas pruebas implican procedimientos invasivos o molestos, por lo que es básico que el paciente esté informado y dispuesto a colaborar.

El objetivo de aplicar una nueva prueba diagnóstica es obtener información diferente a la ya disponible que sea beneficiosa para mi paciente. La utilidad de una prueba es indiscutible cuando la enfermedad, si no se diagnostica, comporta un peligro para la vida del paciente, la prueba no produce efectos adversos importantes y existe un tratamiento efectivo para esa enfermedad.

Bibliografía:

- 1.- Ramos JM, Hernández I. Investigación sobre pruebas diagnósticas en Medicina Clínica. Valoración de la metodología. *Med Clin (Barc)*. 1998;111:129-34.
- 2.- Pozo F. La eficacia de las pruebas diagnósticas (I). *Med Clin (Barc)*. 1988;90:779-85.
- 3.- Pozo F. La eficacia de las pruebas diagnósticas (II). *Med Clin (Barc)*. 1988;91:177-83.
- 4.- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-30.
- 5.- 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.
- 6.- Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res*. 1999;8:113-34.
- 7.- Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. *Stat Methods Med Res*. 1998;7:337-53.
- 8.- Cabello López JB, Pozo Rodríguez F. Estudios de evaluación de las pruebas diagnósticas en cardiología. *Rev Esp Cardiol*. 1997;50:507-19.
- 9.- Pauker SG, Kasirer JP. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302:1109-11.
- 10.- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic test. *N Engl J Med*. 1978;299: 926-30.
- 11.- Ochoa Sangrador C. Valoración crítica de documentos científicos. Aplicabilidad de los resultados de la valoración a nuestra práctica clínica. *Bol Pediatr*. 2002;42:120-30.
- 12.- Ochoa Sangrador C, Orejas G. Epidemiología y metodología científica aplicada a la pediatría (IV): Pruebas diagnósticas. *An Esp Pediatr*. 1999;50:301-14.
- 13.- 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;271:389-91.
- 14.- Jaeschke R, Guyatt GH, 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.
- 15.- Department of Clinical Epidemiology and Biostatistics. McMaster University Health Science Center. How to read clinical journals: II. To learn about a diagnostic test. *Can Med Assoc J*. 1981;124:703-10.
- 16.- González de Dios J. Revisión crítica de publicaciones sobre pruebas diagnósticas. *Pediatr Integral*. 2002;6:551-6.
- 17.- González de Dios J. Aplicabilidad de los resultados de nuestra valoración crítica de documentos a nuestra práctica clínica. *Pediatr Integral* 2003;7:238-40.

Figura 1. Esquema de los estudios de evaluación de pruebas diagnósticas



- (1) patrón de referencia (*gold standard*) válido
- (2) comparación independiente
- (3) espectro de pacientes adecuado
- (4) descripción completa de los métodos
- (5) presentación correcta de los resultados
- (6) repercusión sobre el manejo diagnóstico-terapéutico

Ppre = probabilidad preprueba; Ppost = probabilidad posprueba

TABLA 1: Sesgos potenciales de los estudios sobre pruebas diagnósticas *

Sesgo de selección:	
MP:	no se tiene en cuenta el espectro clínico de la enfermedad o comorbilidad.
C:	sobreestima S y E si se representan los casos graves; infraestima S y E si se representan los casos leve.
MC:	representar el espectro completo en la muestra; describir el espectro en el análisis; análisis de la prueba diagnóstica en los subgrupos.
Sesgo de la prueba de referencia imperfecta:	
MP:	no se dispone de una buena prueba de referencia y se usa la disponible (aunque no clasifique bien y discrimine correctamente sanos y enfermos).
C:	generalmente sobreestima S y E, pero a veces los infraestima.
MC:	seguimiento clínicos de los pacientes para conocer si son enfermos o no; correcciones matemáticas si se dispone de un subconjunto de pacientes con una adscripción definitiva.
Sesgo de incorporación:	
MP:	elementos de la prueba diagnóstica forma parte de la prueba de referencia.
C:	sobreestima S y E.
MC:	conceptualización adecuada de la prueba diagnóstica y de la prueba de Referencia.
Sesgo de revisión o valoración ciega:	
MP:	la interpretación de la prueba diagnóstica y la prueba de referencia se realiza conociendo el otro resultado, es decir, de modo no ciego.
C:	sobreestima S y E.
MC:	enmascaramiento de las personas que interpretan (y realizan) la prueba diagnóstica y la prueba de referencia.
Sesgo de secuencia o verificación diagnóstica:	
MP:	el resultado de la prueba diagnóstica condiciona la realización de una prueba de referencia.
C:	sobreestima S e infraestima E.
MC:	realización estándar en todos los pacientes del estudio; si no es posible hacerlo: seguimiento de las pruebas negativas y correcciones matemáticas.
Sesgo por resultados no interpretables:	
MP:	es una eventualidad que se produce en cualquier prueba diagnóstica o de referencia.
C:	sobreestima S y E.
MC:	repetición de la prueba si es posible; Inclusión en el estudio de los casos no interpretables.
Sesgo por variabilidad en la interpretación de resultados:	
MP:	diversos observadores que actúan dentro del estudio tienen diferentes S y E; el mismo observador cambia su S y E dentro del estudio debido al entrenamiento.
C:	generalmente infraestima S y E.
MC:	estudios previos (piloto) de consistencia inter e intraobservadores; correcciones matemáticas.
MP: modo de producción; C: consecuencias; MC: modo de control S: sensibilidad; E: especificidad	
* Tomado de ref. nº 8	

Tabla 2. Preguntas para valoración crítica de artículos sobre diagnóstico*	
A.- ¿Son válidos los resultados del estudio?	
Preguntas de eliminación:	
1.	1.- ¿Existió una comparación con una prueba de referencia adecuada?
2.	2.- ¿Incluyó la muestra un espectro adecuado de pacientes?
3.	3.- ¿Existe una adecuada descripción de la prueba?
Preguntas de detalle	
1.	4.- ¿Hubo evaluación "ciega" de los resultados?
2.	5.- ¿La decisión de realizar el patrón de oro fue independiente del resultado de la prueba problema?
B.- ¿Cuáles son los resultados?	
1.	6.- ¿Se pueden calcular los cocientes de probabilidad (likelihood ratios)?
2.	7.- ¿Cuál es la precisión de los resultados?
C.- ¿Son los resultados aplicables al escenario?	
1.	8.- ¿Serán satisfactorios en el ámbito del escenario la reproductibilidad de la prueba y su interpretación?
2.	9.- ¿Es aceptable la prueba en este caso?
3.	10.- ¿Modificarán los resultados de la prueba la decisión sobre cómo actuar?
(*) Según orientación del programa CASPe. En la mayoría de las preguntas se pueden dar tres tipos de respuesta: Si, No, No se puede saber.	

Tabla 3. Interpretación de los cocientes de probabilidad	
CP+ >10 ó CP- <0,1	generan cambios amplios y a menudo concluyentes desde una probabilidad preprueba hasta una probabilidad postprueba
CP+ 5 - 10 y CP - 0,1-0,2	generan cambios moderados desde la probabilidad preprueba hasta la probabilidad postprueba
CP+ 2 - 5 y CP- 0,5 - 0,2	generan cambios pequeños (pero en ocasiones importantes) de la probabilidad
CP+ 1 - 2 y CP- 0,5 - 1	alteran la probabilidad en un grado insignificante (y rara vez importante)
CP+: cociente de probabilidad positivo; CP-: cociente de probabilidad negativo	
* Según el Evidence-Based Medicine Working Group	

Tabla 4: Medidas de utilidad de una prueba diagnóstica (Resultados de un estudio hipotético de la utilidad de una prueba diagnóstica, utilizando su patrón de referencia)			
Prueba diagnóstica	Prueba de referencia		
	Positiva	Negativa	
Positiva	225 VP(a)	25 FP(b)	250 a+b
Negativa	100 FN(c)	400 VN(d)	500 c+d
	325 a+c	425 b+d	a+b+c+d 750
a) Verdaderos positivos (VP): enfermos con la prueba positiva (b) Falsos positivos (FP): no enfermos con la prueba positiva (c) Falsos negativos (FN): enfermos con la prueba negativa (d) Verdaderos negativos (VN): no enfermos con la prueba negativa a+c Casos con patrón de referencia positivo (enfermos) b+d Casos con patrón de referencia negativo (no enfermos) a+b Casos con la prueba diagnóstica positiva c+d Casos con la prueba diagnóstica negativa Sensibilidad (S) = a/(a+c) = 225/325 = 69,2% Especificidad (E) = d/(b+d) = 400/425 = 94,1% Falsos negativos (FN) = 1-S = 30,8% Falsos positivos (FP) = 1-E = 5,9% Valor predictivo positivo (VPP) = a/(a+b) = 225/250 = 90,0% Valor predictivo negativo (VPN) = d/(c+d) = 400/500 = 80,0% Cociente de probabilidad positivo (CP+) = S/(1-E) = 69,2/5,9 = 11,7% Cociente de probabilidad negativo (CP-) = 1-S/E = 30,8/94,1 = 0,32% Prevalencia o probabilidad preprueba (Ppre) = (a+c)/(a+b+c+d) = 325/750 = 43,03% Odds preprueba = Ppre/(1-Ppre) = 43,3/56,7 = 0,76 Odds postprueba = CP+ x odds preprueba = 11,7 x 0,76 = 8,89 Probabilidad postprueba (Ppost) = odds postprueba/(odds postprueba + 1) = 8,89/9,89 = 89,8%			

Figura 2. Nomograma de Fagan

