

¿Cómo leer un artículo de pruebas diagnósticas?

Esta nueva entrega está dedicada a intentar arrojar luz sobre un tema que para los médicos es muchas veces oscuro: ¿cómo se evalúa una prueba diagnóstica?

Esta guía puede ser utilizada para saber si el artículo de algún test diagnóstico que tiene frente a sus ojos, es un artículo que puede tener utilidad o no para ser utilizado en su o sus pacientes.

Los principios de la guía se aplican de igual modo tanto a los artículos que uno busca (en general en Medline) para responder a la pregunta en base un paciente, como a un artículo que cae en nuestras manos por otras razones. En la sección EOPS de detallan y ejemplifican estos pasos iniciales a partir de pacientes reales.

A continuación iremos viendo paso a paso si podemos aplicar los resultados de ese artículo en nuestro paciente.

Introducción

Las pruebas diagnósticas, también llamados exámenes diagnósticos o tests diagnósticos, son ubicuos en la práctica cotidiana del médico de cualquier especialidad.

Al solicitar o interpretar una prueba diagnóstica, nos enfrentamos muchas veces a preguntas que nos cuesta resolver. Con la proliferación de la tecnología y la aparición de nuevos tests, la correcta interpretación de artículos sobre los mismos se hace más importante.

Una vez que decidimos que un artículo es potencialmente relevante (o que el título y el resumen sugieren que la información puede ser importante para el paciente), se plantean las mismas tres preguntas generales que ya comentamos la entrega pasada (ver tabla 1).

Tabla 1.

A. ¿Son válidos los resultados del estudio?

Criterios principales

¿Hubo una comparación ciega, independiente, con un test de referencia, en todos los pacientes?

¿Hubo un espectro adecuado de pacientes en los que en la práctica clínica se aplicará el examen diagnóstico?

¿Se aplicó el test de referencia a todos los pacientes, sin importar el resultado del test en estudio?

Criterios secundarios

¿Se describieron los métodos con el suficiente detalle como para permitir su reproducción?

B. ¿Es la evidencia importante?

¿Se presentan los cocientes de probabilidad (likelihood ratios) o se pueden calcular de los datos del estudio?

C. ¿Se puede aplicar esta evidencia válida e importantes al cuidado de mi paciente?

¿Está el test disponible, accesible, reproducible y fácilmente interpretable?

¿Son aplicables los resultados a mi paciente?

¿Los resultados modificarán mi tratamiento?

¿Se beneficiará mi paciente como consecuencia de la prueba diagnóstica?

A. ¿Son válidos los resultados del estudio?

Para poder creerle a un estudio, tenemos que ver los métodos utilizados por el mismo. Un estudio es válido cuando sus resultados se acercan a la verdad. En este caso que las características del test diagnóstico están lo suficientemente cercas de la verdad como para que valga la pena seguir analizando el estudio en profundidad.

Estos son los puntos más importantes para poder "creerle" a los resultados de un estudio de una prueba diagnóstica.

Criterios primarios

¿Hubo una comparación ciega, independiente, con un test de referencia, en todos los pacientes?

Para evaluar cómo funciona un test diagnóstico es necesario compararlo con la "verdad". Esta verdad está representada por la definición de la enfermedad al momento de realizar el estudio y debe ser lo más objetiva posible (algunos ejemplos son la coronariografía para enfermedad coronaria, la anatomía patológica para un tumor, etc.). A este patrón de referencia se lo denomina **test de referencia o "gold standard"**. Si el test de referencia utilizado no es adecuado, es poco probable que el artículo proporcione resultados válidos.

En el caso de que este sea adecuado, y que se hayan hecho ambos tests a todos los pacientes, la siguiente pregunta es si los resultados de ambos (el que se está evaluando y el de referencia) fueron evaluados de manera independiente y ciega entre sí. Esto significa que la persona que interpretaba uno no tenía conocimiento del resultado del otro test. La importancia de este punto es fácil de entender: es muy común que al ver un nódulo de pulmón en una tomografía, volvemos a la radiografía y advertimos la lesión previamente no detectada; o luego de obtener los resultados de un ecocardiograma auscultamos un soplo previamente inaudible. Cuanto más probable sea que los resultados de un test puedan influir en la interpretación del otro, más importante es la interpretación independiente de ambos.

Por último, es importante que todos los pacientes, sin importar el resultado del test a evaluar, se les realice el test de referencia. Si esto no se realiza lo se pueden estimar correctamente las características del test a evaluar.

¿Hubo un espectro adecuado de pacientes en los que en la práctica clínica se aplicará el examen diagnóstico?

Esto apunta a evaluar si los pacientes del estudio son una población a la que habitualmente se solicitaría el estudio en la práctica cotidiana. Es decir, no solo los pacientes con la enfermedad, sino otros trastornos que pueden confundirse con la misma. También es importante que incluya pacientes con distintos estadios de la enfermedad; ya que si se aplica el test solo a los muy enfermos o a los muy sanos, los resultados no serán aplicables a la totalidad.

Un ejemplo de esto es la utilidad del antígeno carcinoembrionario (ACE) para la detección de cáncer de colon. Los estudios iniciales con resultados muy buenos incluyeron pacientes con carcinoma avanzado y pacientes sanos. Cuando se hicieron estudios con pacientes con cáncer en distintos estadios, las características del ACE empeoraban notablemente.

A veces los investigadores se timentan a no hacer el test de referencia, muchas veces más invasivo que el test en estudio. En algunos casos es lícito usar otros criterios en los pacientes que tienen el test en estudio negativo, cuando el test de referencia es invasivo. Por ejemplo, en un estudio para diagnóstico de tromboembolismo pulmonar, se siguió a los pacientes que no se habían hecho la angiografía pulmonar (test de referencia) por un año. Si en ese año no desarrollaban episodio embólico alguno, se los asumía como que no habían tenido TEP.

Criterios secundarios

¿Se describieron los métodos con el suficiente detalle como para ser reproducible?

Si está evaluando un test que se piensa utilizar en sus pacientes, es importante que exista una descripción detallada de las condiciones en las que se realizó el test. Estas incluyen tanto la preparación del paciente (alimentación, medicaciones a evitar, precauciones a tomar luego del estudio), como la realización de la prueba y el análisis e interpretación de los resultados.

Si llegó a esta altura y se cumplen los criterios mencionados se puede concluir que los resultados del estudio probablemente representen una estimación no sesgada de las características del test. El próximo paso es saber cuáles son las características del test y su utilidad ante una situación particular.

B. ¿Es la evidencia importante?

¿Se presentan los cocientes de probabilidad (likelihood ratios) o se pueden calcular de los datos?

Probabilidades Pre-test y Post-test

Es importante saber que los tests no son perfectos, por lo tanto un test positivo no implica necesariamente la presencia de enfermedad, así como uno negativo no la descarta. Lo que hacen los tests es modificar la probabilidad del paciente de padecer la enfermedad. Pero, ¿cuál es esa probabilidad de padecer la enfermedad? Esta dependerá del cuadro clínico del paciente y del grado de sospecha de la enfermedad. A esta probabilidad se la denomina **probabilidad previa o pre-test**: es la probabilidad que le asignamos al paciente de tener determinada enfermedad antes de solicitar un test (ver editorial de probabilidad).

Entonces lo que va a hacer un test es alterar las probabilidades previas. Así, luego de realizarse el test el paciente tendrá otra probabilidad de enfermedad (**probabilidad posterior o post-test**). Cuanto mejores sean las propiedades de un test más se va a modificar la probabilidad previa.

En realidad, cada dato de la historia clínica o el examen físico va alterando la probabilidad de enfermedad. Un paciente de 50 años, hipertenso, que nos describe un dolor de pecho asociado al ejercicio, va a tener una probabilidad muy distinta de tener enfermedad coronaria que una mujer de 30 años sin factores de riesgo y con un dolor no relacionado al esfuerzo. Esta probabilidad pre-test ejerce una influencia primordial en el proceso diagnóstico.

Las propiedades del test se pueden expresar con distintas características. De ellas, la que mejor logra condensar la utilidad y el poder de un test es el **cociente de probabilidad**.

Cociente de probabilidad

Es la mejor medida para evaluar la utilidad del test, la precisión con la que identifica determinada enfermedad. Nos da una idea del poder de ese resultado para modificar la probabilidad pre-test.

Ilustraremos esto con un ejemplo. Tomaremos un ejemplo clásico de la literatura, el estudio PIOPED, que evaluó el centellograma ventilación/perfusión (CVQ) contra la angiografía para el diagnóstico de tromboembolismo pulmonar (TEP). Hubo 251 pacientes con TEP demostrado por angiografía, y 630 en los que se excluyó el TEP. El CVQ fue clasificado en cuatro categorías, desde una probabilidad elevada hasta normal o casi normal.

Tabla 2

Resultado CVQ	Tromboembolismo Pulmonar				Cociente de Probabilidad
	Presente		Ausente		
	N	Proporción	N	Proporción	
Prob. Elevada	102	102/251 = 0.406	14	14/630 = 0.022	18.3
Prob. Intermedia	105	105/251 = 0.418	217	217/630 = 0.344	1.2
Prob. Baja	39	39/251 = 0.155	273	273/630 = 0.433	0.36
Normal/Casi normal	5	5/251 = 0.020	126	126/630 = 0.200	0.10
Total	251	+	630	=	8.84

Podemos ver varias cosas en esta tabla, que nos ayudarán para calcular el CP. Por un lado, ¿Cuán probable es un CVQ con probabilidad elevada entre los enfermos de TEP? En este caso, observamos que 102 de 251 pacientes con TEP (o 0.406) presentaron un CVQ de probabilidad elevada. ¿Qué sucede en el caso de los que no tuvieron TEP? En ellos, sólo 14 de 630 tuvieron un CVQ de probabilidad elevada (0.022). Al cociente de estas dos probabilidades se lo denomina cociente de probabilidad (CP), (likelihood ratio en inglés). Para un resultado del CVQ de alta probabilidad es de $0.406/0.022 = 18.3$. Esto significa que es 18 veces más probable que un resultado del CVQ de alta probabilidad venga de un enfermo en relación a un sano. Lo útil del CP es que puede calcularse para cada nivel de resultado del test (ver tabla 2). Por ejemplo, si el CVQ da normal o casi, el CP se hace dividiendo esos dos cocientes ($5/251$, o los TEP con CVQ normal; y $126/630$, o los no TEP con CVQ normal). Esta división, $0.02/0.2 = 0.1$. ¿Qué significa esto? Nos dice que un resultado normal del CVQ es más probable que provenga de un paciente sin TEP que de uno con TEP. Como vemos, un CP mayor a 1 implica que ese resultado es más probable en pacientes con la enfermedad, y cuanto más alejado de 1, mayor es la probabilidad de observar ese resultado en un paciente enfermo, y más aumentará la probabilidad post-test. A la inversa, si el CP es menor a 1, la probabilidad de enfermedad disminuye; así cuanto más pequeño sea el CP menor será la probabilidad post-test de enfermedad. De esto se desprende que un CP igual a 1 no cambia la probabilidad de enfermedad, o lo que es

lo mismo, la probabilidad post-test será igual a la pre-test.

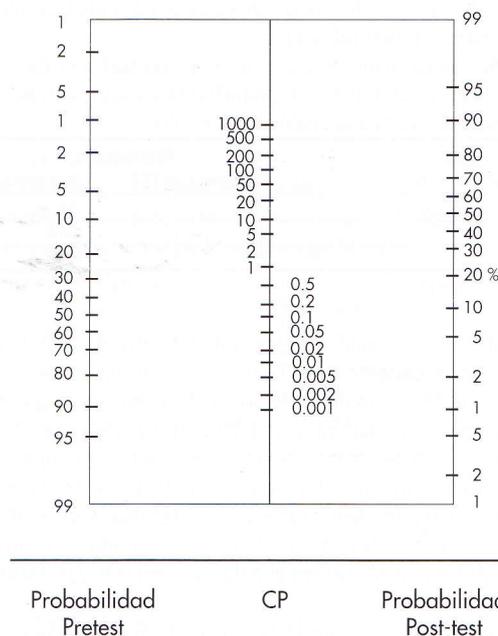
Aunque su utilidad puede variar en distintas situaciones, esto puede ser de utilidad como guía práctica acerca de los CP.:

- Si el CP es mayor a 10 o menor a 0.1, la modificación de la probabilidad pre-test a la post-test suele ser concluyente para confirmar o descartar la enfermedad en cuestión.
- Valores de CP entre 5-10 y entre 0.1-0.2 generan cambios moderados desde la probabilidad pre-test a la post-test.
- CP entre 2-5 y 0.5-0.2 producen cambios pequeños (aunque a veces importantes) de la probabilidad de enfermedad.
- CP entre 1-2 y entre 0.5-1 producen cambios insignificantes de la probabilidad (rara vez importantes)

Ahora que sabemos el significado y la importancia del CP, el próximo paso es evaluar cuánto cambia la probabilidad pre-test de mi paciente luego de un resultado determinado de un test.

Para esto existen dos métodos, el método de los odds y el del normograma. Nosotros nos dedicaremos al del normograma, más fácil y rápido (los interesados en el método de los odds pueden consultar las referencias). El normograma que observan en la figura fue propuesto por Fagan, y permite de una manera sencilla calcular la probabilidad post-test. En la primera columna observamos las probabilidades pre-test posibles. En la segunda columna observamos los distintos posibles CP como resultados del test. En la tercera columna se observan las probabilidades post-test. Así, para saber la probabilidad de nuestro paciente de tener determinada enfermedad, solo debemos colocar una regla que una las primeras dos columnas con los valores adecuados (unir la probabilidad que le asignamos a nuestro paciente con el CP del resultado del test). Siguiendo la línea de la regla, el punto en el que cruce la tercera columna nos arrojará la probabilidad post-test.

Figura 1. Normograma: de la Probabilidad Pre-test a la Probabilidad Post-test



Veamos un ejemplo: tomemos el caso de una paciente de 65 años, en el post-operatorio de una cirugía de cadera que comienza con disnea hace 24 hs. Estaba en tratamiento con heparina profiláctica. De todos modos, supongamos que ud. le asigna, (luego de conocer otros datos de la historia clínica, examen físico Rx Tx y gases), una probabilidad de TEP del 60%. Este valor es la probabilidad pre-test. ¿Qué pasaría con la probabilidad de TEP de esta paciente si el CVQ es de probabilidad alta? Si ponemos una regla que una la probabilidad pre-test, en este caso de 60%, con el CP de ese resultado del test (en este caso de 18.3), observamos que la probabilidad post-test es mayor al 95%. Pero vemos qué pasa si el resultado del CVQ es distinto: si da una probabilidad intermedia (CP1.2), la probabilidad se modifica poco (del 60% cambia a alrededor del 65%); en cambio si el resultado es normal (CP 0.10), la probabilidad post-test desciende a cerca del 13%.

Como la probabilidad pre-test es algo incierta y es una estimación, lo que podemos hacer a continuación es evaluar los posibles resultados dentro de un rango de probabilidades pre-test. En este caso podríamos hacer el mismo ejercicio con probabilidades previas de 50 y 70%, y ver como se modificaría la probabilidad post-test.

Aunque los CP son la mejor manera de expresar la potencia de un resultado de un test para aumentar o disminuir la probabilidad de enfermedad, existen otros parámetros utilizados en la literatura para evaluar las características de un test diagnóstico. Estos son la sensibilidad y la especificidad. Se denomina sensibilidad al porcentaje de individuos enfermos que tienen el test positivo (¿a qué porcentaje de los enfermos detecta el test?). Se denomina especificidad al porcentaje de individuos sin la enfermedad cuyo resultado es negativo. Para calcular estos valores, a través de la tabla de 2x2, debemos dividir los resultados del test en normales o anormales. (ver tabla 3).

Tabla 3

Tabla de 2 x 2

		Test de referencia		
		Enfermedad +	Enfermedad -	
Test	+	a (VP)	b (FP)	VP: Verdaderos Positivos VP: Falsos Positivos VN: Verdaderos Negativos FN: Falsos Negativos
	-	c (FN)	d (VN)	
		Sens (a/a+c)	Espec (d/b+d)	

- **Sensibilidad:** Proporción de pacientes enfermos que tienen el test positivo
- **Especificidad:** Proporción de pacientes sin enfermedad que tienen el test negativo
- **Coefficiente de Probabilidad Positivo:** Sens/(1-Esp)
- **Coefficiente de Probabilidad Negativo:** (1-Sens)/Esp

Si vemos la tabla 2, observamos que es una tabla 4 x 2, no hay un resultado normal o anormal, sino distintos niveles de resultados. Supongamos que solo llamamos anormal o positivo a un CVQ de alta probabilidad. Con los datos de la tabla 2 se puede construir una tabla 2 x 2 para este caso (ver tabla 4)

Tabla 4. Comparación de los resultados del test a evaluar (CVQ) con los del test de referencia (angiografía) tomando como positivos sólo los resultados de alta probabilidad del CVQ.

Categoría de CVQ	Angiografía	
	Presencia de TEP	Ausencia de TEP
Probabilidad elevada	102	14
Otras	149	616
Total	251	630

Sensibilidad, 41%; especificidad, 98%; CP de un resultado de probabilidad elevada, 18.3; CP de otro resultado, 0.61.

Para calcular la sensibilidad a partir de los datos de la tabla 2, vemos a cuántos de los pacientes con TEP tuvieron un resultado de alta probabilidad (102 de 251, o alrededor del 41%). Para obtener la especificidad, vemos cuántos de los pacientes sin TEP (630) tuvieron resultados normales (llamamos normales en este caso a los que no tuvieron probabilidad alta en el CVQ, 616 pacientes). Este resultado (616/630 = 98%) nos da la especificidad. También podemos calcular con esta tabla los CP positivos y negativos con este valor de corte, 18.3 y 0.6 respectivamente (si lo calcula y le da algo distinto puede ser por el redondeo utilizado para facilitar la comprensión).

Pero veamos que pasa si decidimos poner un distinto valor de corte y llamamos normal a un CVQ normal, y anormal a todos los otros resultados (tabla 4). Ahora la sensibilidad asciende a un 98% (246/251), pero cae la especificidad a un 20% (de todos los pacientes sin TEP (630), sólo un 20% (126) presentan resultados normales. Los respectivos CP son 1.23 y 0.1. Tengan en cuenta que usando este punto de corte no sólo perdemos la información diagnóstica de un resultado de probabilidad elevada del CVQ, sino que también interpretamos que un CVQ de probabilidad baja aumenta la probabilidad de TEP, cuando en realidad le disminuye.

Como vemos, utilizando la sensibilidad y especificidad nos perdemos información valiosa, o sino hay que calcular estos valores para cada punto de

corte del test. Esto se utiliza cuando se quiere comparar dos test, se grafican las sensibilidades y especificidades de los distintos puntos de corte para ver cuál tiene mejor área bajo la curva. A este método se lo denomina curvas de la característica operativa del receptor (ROC es su sigla en inglés). Es por ello que el enfoque más utilizado actualmente es el de los CP, debido a su simplicidad y eficiencia.

C. ¿Se puede aplicar esta evidencia válida e importantes al cuidado de mi paciente?

¿Está el test disponible, accesible, reproducible y fácilmente interpretable?

Antes de aplicar el test en su paciente, debe evaluar si es un test reproducible (capacidad para producir un mismo resultado en una misma población). Una baja reproducibilidad se puede deber a la utilización de distintos reactivos, o también a la distinta interpretación de distintos observadores. Es importante que un artículo describa la reproducibilidad del test, especialmente cuanto más subjetividad haya involucrada en su interpretación (pensar por ejemplo en la ecografía, muy observador dependiente). Si la reproducibilidad del estudio es moderada, y es frecuente el desacuerdo clínico, y a pesar de ello el test discrimina bien entre los individuos con y sin el proceso objetivo, estamos frente a una buena señal para utilizarlo en nuestro contexto clínico.

¿Son aplicables los resultados a mi paciente?

Aquí se está preguntando si el examen tendrá la misma precisión en su paciente. Hay que evaluar si el estudio se realizó en un contexto similar al de su práctica. A pesar que no sea un paciente igual al del estudio, ¿es lo suficientemente distinto como para sospechar que los resultados puedan no ser aplicables? (pensar aquí en el estadio de la enfermedad en cuestión, en otros diagnósticos diferenciales contemplados). Si las poblaciones son muy diferentes, los CP no son directamente trasladables a su paciente. Si se encuentra en la literatura una revisión sistemática del test en cuestión, se pueden sacar conclusiones más precisas sobre la generalizabilidad de los resultados.

¿Los resultados modificarán mi tratamiento?

Este punto es importante plantearlo antes de solicitar cualquier test. Se puede hacer el ejercicio de preguntarse: ¿cambiará en algo mi conducta ante un resultado positivo o negativo en este paciente?. ¿Se modificará lo suficiente la probabilidad pre-test ante un resultado positivo o negativo como para modificar su conducta?

En primer lugar, para saber cuánto se modificará la probabilidad de su paciente, es preciso tener una idea de la probabilidad pre-test del mismo. Esto se puede realizar utilizando la experiencia propia, datos estadísticos de prevalencia, ocasionalmente bases de datos, o tal vez de estudios primarios de la literatura que se planteen ese tema.

En general la mayor utilidad de los tests se encuentra cuando la probabilidad pre-test es intermedia. Así, puesto en términos simples, un test positivo hará que se tome una conducta terapéutica debido a la alta probabilidad de enfermedad; y un test negativo bajará la probabilidad lo suficiente como para descartar prácticamente la enfermedad y no administrar tratamiento alguno.

Como vimos anteriormente, el test será de mayor utilidad cuanto más pacientes presenten CP alejados de 1, con lo que modificarán sustancialmente la probabilidad pre-test. En el caso del TEP, en la tabla 2 observamos que el porcentaje de pacientes con resultados extremos del CVQ (probabilidad alta y normal o casi) es de (102 + 14 + 5 + 126)/881 o 247/81 = 28%. Esto refleja el hecho de que a pesar que un resultado extremo puede ayudar mucho, este sólo se observa en una minoría de los pacientes, y muchas veces en la práctica se termina haciendo otro estudio adicional o el test de referencia.

¿Se beneficiará mi paciente como consecuencia de la prueba diagnóstica?

Este es el último paso antes de aplicar un estudio, y sin duda el más importante, el que se refiere a cuál será el beneficio de su paciente con el test. Siempre que el paciente esté dispuesto a realizarlo, el valor de un test preciso será mayor si la enfermedad no diagnosticada es peligrosa, si el test es poco invasivo, y si existe un tratamiento efectivo para la misma.

Dr. Federico Augustovski

Unidad de Medicina Familiar y Preventiva. Hospital Italiano de Buenos Aires.

Bibliografía sugerida

Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine. How to practice & teach EBM. Churchill Livingstone 2000.
 Jaeschke R, Guyatt GH, Sackett DL et al for the Evidence-Based Medicine Working Group. Users' Guides to The Medical Literature III. How to use an article about a diagnostic test. A. Are the results of this study valid? JAMA 1994; 271:389-91.
 Jaeschke R, Guyatt GH, Sackett DL et al for the Evidence-Based Medicine Working Group. 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? JAMA 1994; 271:703-7.