

Cuando la evidencia evalúa pruebas diagnósticas... ¿Qué debemos saber, qué debemos hacer?

Diana Torres, Fernando Sierra, Oscar A. Beltrán

Resumen

El comienzo del uso de la medicina basada en la evidencia (MBE) está cuando reconocemos que en la práctica diaria, para ofrecer mayores beneficios, es necesario recolectar toda la información existente en la literatura que permita acertar en los diagnósticos, procurar la mejor alternativa terapéutica, ofrecer la mejor calidad de vida y pronóstico para nuestro paciente en particular y la sociedad en general. Esta información o evidencia debe ser obtenida eficazmente para lograr la importancia, validez y aplicabilidad conveniente. Para el desarrollo de este tema queremos en primera instancia esbozar las razones por las cuales consideramos de importancia fundamental el conocimiento de la epidemiología básica para el mejor ejercicio médico; en segunda instancia explicaremos algunas definiciones que nos permitan entender y evaluar una prueba diagnóstica y por último daremos las preguntas que debemos tener siempre en nuestra mente cuando estamos evaluando literatura médica que involucra pruebas diagnósticas.

En la práctica clínica el médico debe basar sus actos siempre teniendo tres parámetros de relevancia. El primero consiste en que quien da las pautas y caminos por seguir es el paciente, quien nos proporciona la información necesaria para hacer presunciones diagnósticas; el segundo, es la experiencia y el conocimiento que nos da el ejercicio profesional con el transcurrir de los años y el tercero es seleccionar nuevos conocimientos disponibles en la mejor literatura, para lo cual debemos tener la capacidad de evaluar y determinar la validez y magnitud de cada publicación y decidir si para nuestro paciente en particular es relevante lo descrito o no tiene aplicabilidad.

Palabras claves: pruebas diagnósticas, medicina basada en la evidencia.

Summary

Onset of evidence based medicine is when we admit that in order to offer the best benefits in the quotidian practice it is necessary to recollect all the existent information in the literature to allow us to assert in the diagnosis, to use the best therapeutic alternative and to offer the best quality of life and prognosis to our particular patient and for the society in general. This information or evidence must be obtain efficiently to achieve the importance, validity and convenient applicability. To develop this matter, in first instance, we have to rough out the reasons to consider of fundamental importance, the knowledge of basic epidemiology to improve the medical practice; in second instance, we will give some definitions that allow us to understand and evaluate a diagnostic test and finally we will give the questions we have to keep in mind when we are evaluating medical literature that involves diagnostics tests.

In he clinical practice the physician must base his acts always having three relevant parameters: the first that gives patterns and ways to follow, is the patient who provide us the necessary information to do diagnostic conjectures, the second is the experience and the knowledge that gives the professional practice throughout the years, and the third is to select recent available knowledge in the best literature evaluating the issue validity and decide if is or not applicable in our particular patient.

Key words: diagnostic test, evidence-based medicine.

Muchos médicos dicen creer en la medicina basada en la evidencia (MBE) pero no cuentan con los suficientes elementos de juicio que les permitan seleccionar adecuadamente lo relevante para su paciente en particular. Es por esto que expertos como el doctor David Sackett afirman que si aplicamos la medicina

basada en la evidencia debemos reconocer que es “el uso concienzudo, explícito y juicioso de la evidencia disponible en las búsquedas sistemáticas para tomar decisiones clínicas”. Cuando vamos a usar la literatura disponible no podemos separar las decisiones que tomemos de nuestra competencia como médicos ya que no sólo el desconocimiento del análisis de la literatura es el problema (pués hay literatura “buena” y “mala”), sino la falta de competencia por parte de los médicos que hace que al decir que basan sus actos en la evidencia disponi-

Diana Torres, MD, Gastroenteróloga, Clínica San Pedro Claver, Hospital de Kennedy. **Fernando Sierra A., MD,** Gastroenterólogo- Epidemiólogo Clínico, Fundación Santa Fe de Bogotá. **Oscar Beltrán, MD,** Gastroenterólogo-Epidemiólogo Clínico, Clínica San Pedro Claver. Juan N. Corpas. Bogotá D.C., Colombia.

Rev Colomb Gastroenterol 2004;19:281-285.

ble pueden convertir la MBE en una tiranización al utilizar información que puede no ser aplicable para un determinado caso en particular. Tal como afirma el doctor Sackett, algunos estudios pueden ser inaplicables o inapropiados para nuestros pacientes a pesar de ser muy bien diseñados.

Por otro lado, nunca podemos separar las consecuencias de nuestras determinaciones para el paciente de la experiencia, el conocimiento y la mejor evidencia disponible, teniendo en cuenta que los cambios o decisiones que tomemos pueden alterar el pronóstico y calidad de vida, tanto de forma negativa como positiva. La utilización indiscriminada de pruebas diagnósticas puede, además de alterar la calidad de vida, aumentar los costos para las entidades de salud conduciendo a un detrimento en la calidad de la oferta futura para nuestra población.

La práctica médica debe basarse entonces en la integración de la mejor evidencia disponible con la mayor aplicabilidad para nuestro medio, es decir no olvidar que cuando leemos un artículo puede aparentemente ser muy válido, pero puede estar mostrando solamente una pequeña parte del verdadero escenario.

Para lo anterior, siempre al estar ejecutando el acto médico debemos preguntarnos sobre la literatura si:

- ¿La literatura disponible tiene una evidencia fuerte y actual? (recuerde que la evidencia está en continuo desarrollo y cambia con el tiempo)
- ¿La literatura en la que estamos basando nuestras decisiones es comprensible?
- ¿La información es la más relevante para nuestro paciente en particular?
- Los datos que provee el artículo son asequibles y fáciles de usar?
- ¿Los datos obtenidos del artículo le ofrecen a mi paciente beneficios en calidad de vida y pronóstico?
- ¿Puede el estudio ser aplicable a mi población en particular? (no caer en el error de escribir lo que hay que hacer para que no se haga)
- ¿El problema tratado en el artículo que estoy revisando es también un problema en mi población?
- ¿Cuál es la intervención que está siendo comparada y si es razonable su comparación?
- ¿El desenlace al aplicar la información dada por este artículo es importante para mis pacientes?
- ¿El costo de aplicar lo publicado en este artículo es útil en el contexto de mi población?

Definiciones

Nuestros esfuerzos en la práctica clínica van encaminados a determinar la mejor alternativa terapéutica, haciendo de las pruebas diagnósticas un eslabón determinante que conduce al diagnóstico, pero para cuya interpretación y adecuada utilización debemos

involucrar la historia, las observaciones clínicas, las pruebas de laboratorio o los estudios de imágenes.

Todas las pruebas diagnósticas ayudan al clínico a estimar la probabilidad particular de que se presente una determinada enfermedad; es así como muy pocas veces podemos decir que el diagnóstico sea una certeza basados en la sola historia clínica, sino más bien logramos usualmente afirmar ante una enfermedad que es probable, posible o improbable su presencia y el grado de certeza se logrará entonces después de aplicar una determinada prueba, y su utilidad dependerá de múltiples factores que incluyen el deseo del paciente o familiares por una terapia, los riesgos y costos de la misma prueba, los riesgos, beneficios y costos de la terapia.

Con lo anterior podemos intuir que una prueba diagnóstica es útil si al aplicarla a un paciente dado, aumenta o disminuye la probabilidad de que dicho paciente presente o no un diagnóstico específico. Esto es lo que se ha denominado la razón de probabilidades (*likelihood ratio*), que en lenguaje sencillo es la probabilidad de que un determinado examen de laboratorio cambie la probabilidad de que un paciente tenga una determinada enfermedad, colocada ésta sobre un determinado umbral que depende de la probabilidad determinada por nuestro juicio clínico de que tenga o no una enfermedad específica (Figura 1) y del comportamiento de la prueba diagnóstica en pacientes con y sin la enfermedad.

Para saber si una prueba diagnóstica que se está evaluando en un determinado estudio es útil o no, se determina el desempeño de la prueba comparándola imparcialmente con otra prueba de referencia que debe ser la “*gold standard*”. Basados en la comparación podemos recordar algunas definiciones útiles en el entendimiento de lo anterior como son:

- Sensibilidad: es la proporción de pacientes con la enfermedad que tienen la prueba positiva (positivos/total de pacientes con enfermedad).
- Especificidad: es la proporción de pacientes sin enfermedad que tienen la prueba negativa

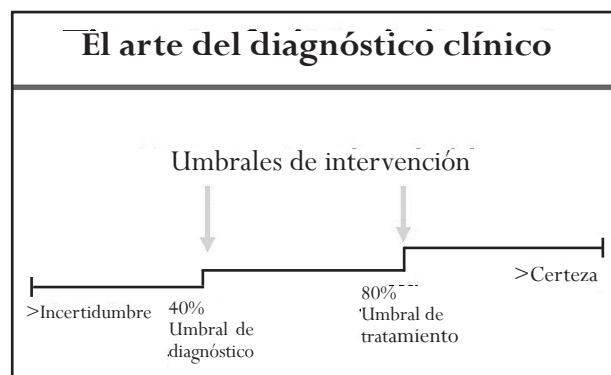


Figura 1. El arte del diagnóstico clínico.

(negativos/ total de pacientes que no tienen la enfermedad).

- Razón de posibilidades, que en los artículos escritos en inglés la encontrarán como *likelihood ratio* o LR: es la probabilidad de que una prueba en particular dé un resultado determinado en pacientes con la enfermedad comparado con la probabilidad de que esa misma prueba dé ese mismo resultado en pacientes que no tienen la enfermedad. Cuando el resultado de la prueba es sólo positivo o negativo, la LR puede ser calculada fácilmente basándose en la sensibilidad y especificidad así:

LR positivo : sensibilidad/1 - especificidad

LR negativo : 1- sensibilidad/especificidad

Cálculo de sensibilidad, especificidad y LR de una prueba diagnóstica:

	Positivos	Negativos
Prueba anormal	a	b
Prueba normal	c	d
	a + c	b + d

Sensibilidad: $a/a+c$

Especificidad: $d/b+d$

LR+ = $a/a+c / 1 - (d/b+d)$

LR- = $1 - (a/a+c) / d/b+d$

Existe otra característica del desempeño operativo de una prueba diagnóstica y son los denominados Valores Predictivos Positivos y Negativos. Pero la interpretación de éstos debe ser muy cuidadosa ya que son altamente sensibles a la prevalencia que tiene la enfermedad en la población donde se desarrolló el estudio clínico para determinarlos. El valor predictivo positivo (VPP) es la probabilidad de que se presente una enfermedad entre quienes tienen una prueba positiva; es usualmente inapropiado a menos que los pacientes a quienes se les va a aplicar una determinada prueba sean similares a los pacientes o población en la cual se calculó el VPP y tengan la misma probabilidad pretest. Este problema no ocurre con otras características operativas de las pruebas ya que la sensibilidad, la especificidad y la LR son similares a través de diferentes poblaciones y son menos dependientes de la prevalencia de la enfermedad en cada población en particular.

Uno de los inconvenientes con la LR, la cual consideramos la mejor herramienta con que contamos los clínicos, es que entre las diferentes características operativas que ofrece una prueba o examen diagnóstico, en sí misma la LR es una *Odds*, y para nosotros no es entendible ni cómodo funcionar en términos de Odds (de hecho no se le encuentra una traducción al español adecuada). Quizás las únicas veces en que nos hemos vistos afrontados al término de Odds en nuestra vida es en el campo deportivo, cuando leemos o escuchamos que las apuestas a favor de un determinado boxeador o de un caballo en una competencia están 8 a 2; esto en otras palabras simplemente es que

por cada 8 apostadores que dicen y apuestan a que determinado boxeador a caballo gane, hay 2 apostadores que apuestan que ese determinado boxeador o caballo no gana. Para hacer más entendible el concepto de odds y su diferencia con el de probabilidad, nos ha resultado muy útil el ejemplo del lanzamiento de un dado; es así como si usted lanza un dado una vez, la probabilidad de que en ese único lanzamiento obtenga un 6 es de 1 en 6, o sea 16,6%, ya que el dado tiene 6 caras. Pero en ese mismo y único lanzamiento la odds de obtener un 6, es 1 contra 5, porque en sólo una vez tendrá la ocasión de obtener el 6 pero a su vez tendrá 5 oportunidades de que en un lanzamiento salga un número diferente de 6.

Para poder obtener la información suministrada por la LR de un determinado examen paraclínico, ésta debe ser multiplicada por la odds de que el paciente tenga la enfermedad antes de practicarle dicho examen, y esa odds la debemos de obtener convirtiéndola desde la probabilidad que nosotros como clínicos le asignamos a nuestro paciente de que tenga o no la enfermedad antes de solicitarle el examen. Como éste es un ejercicio matemático que para muchos clínicos resulta tedioso y complicado, Fagan diseñó una gráfica que es muy fácil de usar (Figura 2) y cuya copia les recomendamos sacar y llevar con ustedes en su tarea clínica diaria. Simplemente, con la ayuda de una pequeña regla ubican la probabilidad de tener la enfermedad que ustedes le han asignado a su paciente en la columna de la izquierda de la tabla de Fagan y pasan la regla para cruzar la LR del examen de laboratorio que viene en la columna intermedia de la tabla y simplemente la extensión de la regla hasta cruzar la columna de la derecha de la tabla les da la probabilidad numérica de que el paciente tenga la enfermedad.

Cabe anotar que muchas pruebas diagnósticas no tienen como alternativa sólo respuestas de resultados positivo o negativo sino que dan rangos; para cada uno de los rangos se puede calcular una determinada LR.

Buscando la mejor evidencia disponible en la literatura

Teniendo en cuenta las bases anteriores y con miras a lograr que el acto médico se base en conocimientos, experiencia y literatura adecuada, Bandler diseñó siete preguntas estratégicas que debemos recordar al analizar o realizar un estudio de pruebas diagnósticas:

- Características de la población: la sensibilidad o especificidad de una prueba depende de las características de la población estudiada; los datos expuestos en una población determinada pueden no ser aplicables a otra población. Para asegurarnos de la aplicabilidad a nuestra población debemos conocer la distribución por edad, sexo,

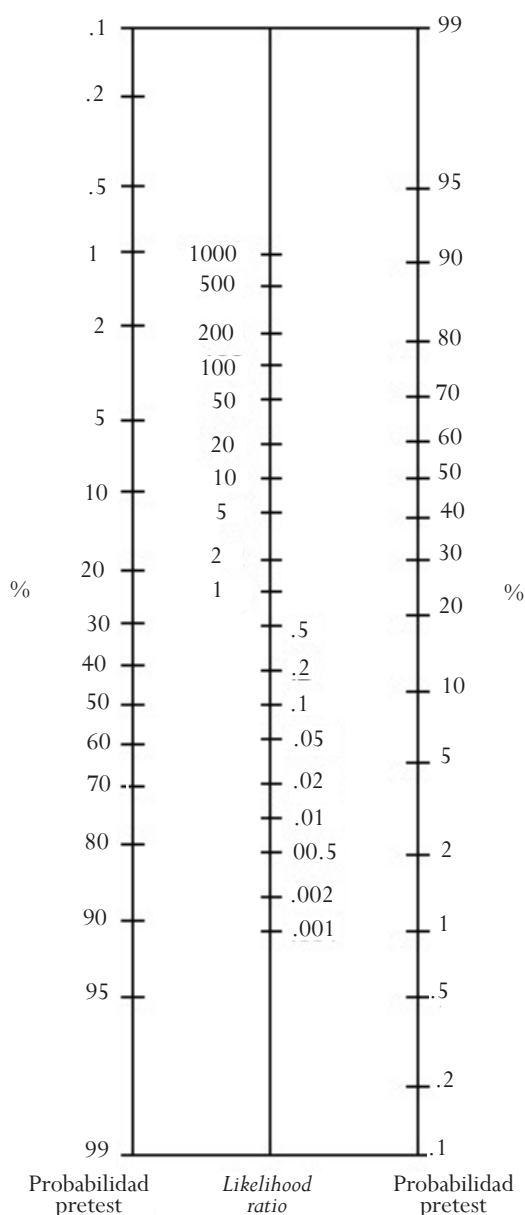


Figura 2. Tabla de Fagan para el cálculo de la probabilidad de enfermedad utilizando el valor de lr del resultado de la prueba de laboratorio.

signos y síntomas iniciales o estadio de la enfermedad, y criterios de elección para los pacientes que entran al estudio.

- Subgrupos adecuados: la sensibilidad y la especificidad pueden estar calculadas para una población muy amplia y al tratar de aplicarlas a subgrupos poblacionales pueden variar. Para que la prueba pueda ser utilizada se deben tener en cuenta diferentes niveles de precisión según los diferentes subgrupos de población. Este se cumple cuando se informa sobre la precisión de la prueba en relación con cualquier subgrupo demográfico o clínico, por ejemplo en pacientes sintomáticos versus asintomáticos.

- Evitar el sesgo de selección: éste se produce cuando la población elegida es derivada de los pacientes con resultado positivo o negativo a una prueba para verificar el diagnóstico mediante otra prueba considerada como "gold standard". Para valorar si este criterio se cumple en los estudios de cohortes, todos los sujetos deben haber sido asignados para recibir tanto la prueba en estudio como el "gold standard", ya sea por un proceso directo o por seguimiento clínico. En los estudios de casos y controles la validez del criterio depende de: si la prueba en estudio es aplicada antes del "gold standard" se cumple este criterio si la verificación de la enfermedad se realiza en la serie consecutiva de los pacientes estudiados sin tener en cuenta los resultados de la prueba en estudio; si la prueba se aplica después del "gold standard" el criterio se cumple si los resultados de la prueba se estratifican de acuerdo con los factores clínicos que pueden estar implicados según el tipo de "gold standard".
- Evitar el sesgo de medición: éste se produce si la prueba en estudio o el "gold standard" se realizan sin tomar precauciones para garantizar la objetividad de su interpretación (similar al ciego en los estudios de tratamiento). Esto se evita si la prueba y el patrón se interpretan de forma separada por personas independientes que desconocen los resultados dados por una y otra. Para que se cumpla este criterio se debe afirmar que los dos test se realizaron en forma independiente.
- Precisión de los resultados: la precisión de la sensibilidad y la especificidad la depende del número de pacientes evaluados. El resultado obtenido debe tener unos intervalos de confianza. Los intervalos de confianza o el error estándar deben estar reflejados independientemente de la magnitud.
- Presentación de resultados indeterminados: no todas las pruebas dan como resultado positivo o negativo, a veces dan uno indeterminado. La frecuencia de resultados indeterminados limita la aplicabilidad de la prueba o la hace muy costosa porque requiere de aplicar otras pruebas diagnósticas para confirmar o no el diagnóstico. Para valorar este criterio el artículo debe informar los resultados positivos, negativos y los indeterminados así como si los resultados indeterminados se incluyeron o no al calcular los indicadores de precisión de la prueba.
- Reproducibilidad de la prueba: se debe valorar si al repetir la prueba da el mismo resultado; la variabilidad del test y la interpretación del observador hacen que los resultados de las pruebas no siempre sean consistentes. Para valorar este criterio en pruebas que requieren interpretación del

observador, alguno de los test debe ser evaluado de alguna forma en que se evite la variabilidad interobservador; para pruebas en que no interviene el observador se debe minimizar la variabilidad del instrumento.

El estudio de pruebas diagnósticas ideal deberá incluir:

1. Una comparación ciega e independiente con un "gold standard" adecuado.
2. Una muestra de pacientes adecuada que pueda representar el grupo de pacientes a los que se les va a aplicar la prueba diagnóstica.
3. La realización del "gold standard" no se vio influenciada con el resultado de la prueba estudiada
4. Población de estudio es amplia, permitiendo con mayor probabilidad hacer comparaciones
5. Los resultados están dados en términos de sensibilidad, especificidad o LR, o el artículo escribe los datos necesarios para calcularlos. Como la LR+ compara la proporción de pacientes con la enfermedad que tiene la prueba positiva (numerador) con la proporción de pacientes sin la enfermedad que tienen la prueba positiva (denominador), si estas son iguales (es decir numerador = denominador), la LR+ será de 1 y concluiríamos que la prueba diagnóstica no sirve para diferenciar entre enfermedad o no enfermedad. Así entre más lejos se esté del 1 sea para abajo o para arriba mejor será la prueba diagnóstica para diferenciar un grupo de otro. La LR es un estimativo de qué tan bien se desempeña una determinada prueba, basado en una muestra de pacientes del estudio. Si el tamaño de la muestra es pequeño, o el número de pacientes en un grupo es pequeño, los valores obtenidos tienen mayor probabilidad de ser inexactos; para esto se da el intervalo de confianza que permite que la imprecisión sea cuantificada: es así como un IC de 95 % indica que los cálculos no son muy precisos y un valor cercano a 1 indica una LR bastante precisa. Por lo tanto si el IC no se da se deberá intuir que

posiblemente es por que es una muestra muy pequeña que da un IC amplio y menos precisión en los resultados

6. Finalmente, debe permitir ayudar a calcular la probabilidad post test. Para esto se debe calcular la probabilidad pre test que está basada en la experiencia clínica, datos de la población local, o de estudios de la probabilidad de presentación de enfermedades. Este estimativo no necesita ser exacto, un rango de posibilidades puede ser útil; posteriormente se deberá aplicar la LR y se obtendrá la probabilidad post test. La importancia de la probabilidad post test es que este cálculo nos permita tomar decisiones terapéuticas y dar información sobre pronóstico.
7. No podemos olvidar que dependiendo del medio en que estemos trabajando debemos preguntarnos si la disponibilidad de una prueba es posible, así como sus costos y sus beneficios.

Para poder utilizar todo lo anterior en la práctica diaria, sería ideal que todos los médicos tuviéramos acceso a Internet permanentemente o al menos pudiéramos tener una "palm pilot", ya que de esta forma podemos tener información acerca de sensibilidad, especificidad y LR de determinadas pruebas, promoviendo una practica más certera y aproximada a los mejores estándares científicos.

Bibliografía

1. **A. Moyer, A Kennedy.** Understanding and using diagnostic test. Clinics in Perinatology. Volume 30. Number 2. June 2003.
2. **Bandolier.** Bias in diagnostic testing. Evidence based thinking about health care. Dec 1999;70-75.
3. **Bandolier.** Evidence and Diagnostics. Evidence based health care. February 2002
4. **Brandt White.** Making evidence based medicine doable in everyday practice. Family practice Management. Vol 11 Number 2 February 2004
5. **Fagan TJ.** Nomograma for Bayes' theorem. *N Engl J Med* 1975; 293:257
6. **Guyatt GH, et al.** Users' guides to the medical literature. How to use an article about therapy or prevention. Are the results of the study valid? *Jama* 1993; 270(21):2598-2601.
7. **Sackett DL, Rosenberg W.** Evidence based medicine: what is and what it isn't. *BMJ* 1996; 312: 71-72
8. **Sierra F, Torres D, Cárdenas A.** The role of Likelihood ratio in Clinical Diagnosis. Clinical Gastroenterology and Hepatology. In press.