

# Medicina Basada en la Evidencia (2da Parte) - Estudios de diagnóstico

Dr. Hernán Rowensztein  
Dra. Valeria Berlín.\*

**E**n esta ocasión nos dedicaremos a estudiar cómo se lee un estudio sobre métodos diagnósticos. Como vimos en la edición sobre Estudios de Intervención, seguiremos la sistemática recomendada por las guías para los usuarios de la literatura científica del Journal of the American Medical Association<sup>1</sup>. Para analizar la validez del estudio analizaremos la validez interna, la magnitud de los resultados y luego la validez externa<sup>2</sup>. Utilizaremos como ejemplo un artículo publicado en la revista Pediatrics<sup>3</sup>, que fue elegido por su sencillez, y estudia la validez de los métodos rápidos para la detección de la infección urinaria en niños.

*Estudios sobre pruebas o tests diagnósticos: El diseño ideal para validar un método diagnóstico debe ser:*

1. **Comparativo:** se compara la capacidad diagnóstica del método en estudio contra un test de referencia. A diferencia de lo que vimos en los estudios de intervención, donde la comparación se efectúa entre datos de dos conjuntos de sujetos (comparación intergrupo o intersujeto), en este caso la comparación se denomina intrasujeto o intragrupo porque el resultado de la prueba en cada paciente se compara contra el resultado del patrón de referencia en el mismo sujeto.
2. **Transversal:** cada prueba se realiza una sola vez en cada paciente a diferencia de los estudios longitudinales en los cuales se realizan mediciones en forma seriada.
3. **Observacional:** el investigador no realiza ninguna intervención sobre los sujetos, solo observa.

**La validez de la prueba Uriscreeen para la detección precoz de la infección urinaria en niños. Yehezkel Waisman, MD; Zerem Elisheva, MD; Lisa Amir, MD, MPH, y Marc Mimouni, MD**

**Resumen.** Objetivo: Para determinar la validez del Uriscreeen, una prueba rápida de diagnóstico basado en la detección de la catalasa en orina para la detección temprana de orina infección del tracto (IU) en niños, en comparación con el análisis del sedimento de orina y pruebas de varilla (Multistix).

Diseño del estudio: Estudio transversal. Niños de 1 mes a 17 años de edad que acudieron al Servicio de Urgencias de un centro pediátrico de atención terciaria, entre marzo y noviembre de 1996 con síntomas sugestivos de infección urinaria. Métodos. Las muestras de orina obtenidas de forma aleatoria sobre una muestra de 121 pacientes fueron evaluadas de forma simultánea por Uriscreeen (prueba de catalasa), análisis de orina (piuria microscópica), varilla (esterasa leucocitaria y nitritos), y cultivo de orina cuantitativo. Todos las muestras fueron recogidas por una de las tres técnicas estériles (mitad de la micción, sondaje vesical o aspiración suprapúbica), según corresponda a la edad. Utilizando el cultivo de orina cuantitativo como el estándar de oro, se determinaron y compararon la sensibilidad, especificidad y valores predictivos positivos y negativos de todas las pruebas. Se registraron para cada participante edad, sexo, temperatura, y síntomas de presentación. Resultados. De los 121 pacientes, 35 (28,9%) fueron positivos los resultados del cultivo: 30 niñas (85,7%) y 5 varones (14,3%). En comparación con los análisis de orina y pruebas de varilla, el Uriscreeen presentó la mayor sensibilidad (100% vs 88,6% y 97,1%, respectivamente) y el más alto valor predictivo negativo (100% vs 95% y 98,6%, respectivamente), pero las más pobres especificidad (68,6% vs 88,4% y 82,5%, respectivamente) y valor predictivo positivo (56,4% vs 75,6% y 69,4%, respectivamente). Conclusiones: El uso de Uriscreeen para el presunto diagnóstico de infección urinaria en niños es limitada y no es significativamente superior al análisis de orina o la prueba de tira reactiva. Sin embargo, debido a su sensibilidad del 100% y el valor predictivo negativo y su facilidad de uso, rapidez y bajo costo, es altamente recomendable para descartar el diagnóstico de infección urinaria. En los laboratorios, un resultado negativo de Uriscreeen puede prevenir la necesidad de realizar costosos urocultivos. Pediatrics 1999; 104 (4). URL: <http://www.pediatrics.org/cgi/content/full/104/4/e41>

\*Clínica Pediátrica. Hospital de Pediatría Juan P. Garrahan. Buenos Aires, Argentina

4. **Prospectivo:** las mediciones se realizan después de planificar el estudio cuidadosamente, a diferencia de los estudios retrospectivos en los cuales las mediciones ya fueron realizadas antes de la planificación.

El estudio que analizamos hoy aclara que fue realizado en forma prospectiva, transversal y se trató de un estudio observacional.

## Inicio de la lectura crítica

Siempre es conveniente identificar los cuatro componentes del estudio (“PICO”): el paciente (población), la intervención (la prueba diagnóstica), el comparador (el test de referencia) y el “outcome” o resultado (el evento o enfermedad que se quiere diagnosticar).

*En este caso se trata un estudio sobre niños con fiebre sin foco aparente, con o sin síntomas de disuria, dolor abdominal, vómitos o poliuria. Se comparó el urocultivo contra los test rápidos (uriscreeen, varilla multiple y test de esterasa leucocitaria) para la detección de la infección urinaria.*

## Validez interna

### 1.1 Criterios primarios:

(a) ¿Existió una comparación ciega e independiente con un estándar de referencia?

(b) ¿Se evaluó la prueba en un espectro apropiado de pacientes similares a aquellos en los que se aplicará en la práctica?

a) Esto se refiere a si la prueba que estamos evaluando fue comparada con “la verdad”, es fundamental ya que el estándar de referencia nos servirá para asumir si los resultados del test evaluado son ciertos o no. Por lo tanto el estándar de referencia o “gold standard” debe ser cuidadosamente elegido, ya que si no es así podremos cometer un error sistemático al tomar como verdadero un resultado que no lo es o viceversa<sup>4</sup>. Algunos ejemplos son: la hemodinamia para el tromboembolismo pulmonar, la anatomía patológica en un tumor, la biopsia de intestino para la enfermedad celíaca, etc.

En general los estándares de referencia son estudios caros, invasivos, o difíciles de realizar, por este motivo es que se desarrollan test más baratos, menos cruentos, etc. Estas características del estándar de referencia, por si solas no pueden justificar la elección de otro test más sencillo pero menos acertado.

Otras veces no existe un solo “gold estándar”, y para confirmar los resultados se debe recurrir a una batería de tests, o incluso a un seguimiento lo suficientemente prolongado del paciente, como para asegurarse la presencia o ausencia del evento o enfermedad que se quiere diagnosticar.

Luego veremos si se realizó una comparación ciega e independiente, es decir ambos test deben ser realizados por operadores diferentes y desconociendo el resultado de la otra prueba, para de esta forma evitar el sesgo y la subjetividad en las mediciones que provoca el conocer un resultado. Un ejemplo de esto podría ser que

### Guía del usuario para artículos sobre diagnóstico. JAMA 1994; 271: 389-391, 703-707

#### 1. Validez interna

##### 1.1 Criterios primarios:

(a) ¿Existió una comparación ciega e independiente con un estándar de referencia?

(b) ¿Se evaluó la prueba en un espectro apropiado de pacientes similares a aquellos en los que se aplicará en la práctica?

##### 1.2 Criterios secundarios:

(a) ¿Los resultados de la prueba en estudio influyeron en la decisión de realizar el método de referencia (“gold standard”)?

(b) ¿Se describieron los métodos para llevar a cabo el examen con el suficiente detalle como para permitir su reproducción?

#### 2. Análisis de resultados

(a) ¿Cuáles son los coeficientes de probabilidad (LR) asociados a diferentes intervalos de resultados de la prueba?

(b) ¿Cuál es la precisión de las estimaciones?

#### 3. Validez externa: Utilidad de los resultados para la asistencia de mis pacientes

(a) ¿Los resultados son aplicables a los pacientes en mi práctica?

(b) ¿Será posible realizar e interpretar la prueba en forma satisfactoria en mi medio?

(c) ¿Los resultados van a cambiar mi estrategia de tratamiento?

(d) ¿Los pacientes van a estar mejor como resultado de la prueba?

luego de realizar una hemodinamia y confirmar una cardiopatía, será más probable detectar en la ecocardiografía una patología cardíaca que antes pudo haber pasado desapercibida.

*En este caso compararon los test rápidos contra el estándar de referencia que es el urocultivo tomado por medio de la técnica correcta según la edad del niño. No aclara si los test fueron realizados por operadores diferentes y si estos desconocían el otro resultado, pero da a entender que el urocultivo fue realizado en otro sector hospitalario diferente a las tiritas reactivas, y se demoró el resultado de forma que al realizar los métodos rápidos se desconocía el resultado del estándar de referencia.*

**b)** En esta parte se evalúa el tipo de pacientes que fueron incluidos en el estudio. Resulta obvio que sería muy fácil el diagnóstico de una patología en estadios muy avanzados de enfermedad, por lo que es importante que el espectro de pacientes incluidos en el estudio sea lo más amplio posible, idealmente desde personas sanas hasta los muy comprometidos. Si no tenemos esto en cuenta corremos el riesgo de sobreestimar o subestimar el poder diagnóstico del test. Un ejemplo de esto podría ser la observación de piel y mucosas para el diagnóstico de anemia, si tomamos pacientes con anemia profunda (menos de 7 gr%) probablemente diremos que la observación de mucosas es muy adecuada como método diagnóstico. En cambio si el espectro de pacientes incluye algunos con hemoglobinas superiores a 10 gr% probablemente nos equivoquemos y tomemos como sanos a algunos pacientes anémicos.

En nuestro estudio se incluyó a una amplia población, desde lactantes con fiebre sin foco hasta adolescentes con signos y síntomas de compromiso del tracto urinario.

## 1.2 Criterios secundarios:

*(a) ¿Los resultados de la prueba en estudio influyeron en la decisión de realizar el método de referencia (“gold standard”)?*

*(b) ¿Se describieron los métodos para llevar a cabo el examen con el suficiente detalle como para permitir su reproducción?*

Estos criterios (secundarios) aportan fortaleza al estudio en el caso que sean tenidos en cuenta, son complementarios a los criterios primarios pero no son descalificadores en el caso de no ser cumplidos.

**a)** En la práctica médica no se comienza a estudiar a los pacientes con los métodos más caros e invasivos, excepto claro que la sospecha clínica o la gravedad del cuadro lo justifiquen. En el caso de los estudios diagnósticos se debe corroborar que ambas pruebas diagnósticas hayan sido realizadas sin importar el resultado de la primera. Es decir, se debe evitar el sesgo de verificación que se podría producir en el caso que sólo si la primera prueba realizada da positiva se realice la otra<sup>5</sup>.

*En nuestro estudio se realizaron las 4 pruebas diagnósticas a todos los pacientes, sin importar el resultado de todas ellas.*

**b)** En el apartado de Materiales y Métodos del artículo deben estar detallados todos los aspectos técnicos del estudio y el equipamiento. Que estén todos los elementos para que el estudio sea replicable es fundamental. Por ejemplo en un estudio sobre la utilidad de la ecografía en el diagnóstico de una cardiopatía, deben constar el modelo del equipo, el transductor utilizado, la posición del paciente, el entrenamiento del personal que realiza el estudio, etc.

En el estudio analizado se detalla la técnica de toma de la muestra, marcas y método con que se realizaron tanto las pruebas rápidas como el urocultivo. Detalla a que resultado se lo consideró positivo y a cual negativo.

## Análisis de los resultados: ¿son los resultados importantes?

*(a) ¿Cuáles son los coeficientes de probabilidad (LR) asociados a diferentes intervalos de resultados de la prueba?*

*(b) ¿Cuál es la precisión de las estimaciones?*

**a)** Antes de someter a un paciente a un método diagnóstico partimos de una presunción de

enfermedad, lo cual nos lleva en algunos casos a decidir si estudiarlo. Por lo tanto digamos que los médicos tenemos una “probabilidad pre-test” acerca de si el paciente está enfermo o no. Los métodos diagnósticos no son infalibles ni perfectos, y eso lo sabemos bien; por lo que normalmente no confirmamos ni descartamos enfermedad tan solo con un resultado, lo que hacemos es cambiar las probabilidades pre-test del paciente a una “probabilidad post-test” de acuerdo a si el resultado es positivo o negativo, lo que simplemente lo acercará o alejará de la posibilidad de estar enfermo. La magnitud de ese cambio dependerá en gran medida de la “capacidad operativa” del test, no es lo mismo una prueba de PPD negativa en un paciente con sospecha de tuberculosis que un cultivo de esputo definitivamente negativo.

Los resultados de los estudios de diagnóstico pueden presentarse de dos formas: como variables numéricas continuas en el caso que el resultado sea una medida numérica que surge de la media, mediana, etc.; o como una variable dicotómica en el caso que el resultado sea expresado según la capacidad del test de detectar enfermos y sanos, prueba positiva o prueba negativa.

Si tomamos como ejemplo un nuevo equipo para medir la tensión arterial en forma automatizada, un ejemplo del primer caso de medición sería expresar la tensión arterial como la media de los pacientes tomada con el nuevo método y con el estándar de oro (ej: tensión arterial invasiva). Un ejemplo del segundo caso sería, por ejemplo, expresar la capacidad de un nuevo método serológico para detectar a los enfermos por el virus de HIV, el resultado será positivo o negativo y luego se lo compara con el estándar de oro.

Esta segunda forma de expresar los resultados nos permite representar más claramente para el lector la capacidad del método diagnóstico, al poder calcular los cocientes de probabilidad<sup>6</sup>.

Es importante destacar que se puede “convertir” cualquier resultado numérico continuo en dicotómico, simplemente tomando puntos de corte clínicamente significativos. Con los datos divididos en categorías, podemos entonces armar una tabla de 2 x 2:

**Tabla de 2x2 para el cálculo de la capacidad operativa de la prueba diagnóstica.**

|                 | Prueba de referencia positiva | Prueba de referencia negativa |
|-----------------|-------------------------------|-------------------------------|
| Prueba positiva | Verdaderos positivos (VP)     | Falsos positivos (FP)         |
| Prueba negativa | Falsos negativos (FN)         | Verdaderos negativos (VN)     |
|                 | Enfermos                      | Sanos                         |

### Glosario de términos

- **Sensibilidad (S) o tasa de verdaderos positivos (TVP):** La sensibilidad de un método diagnóstico es la capacidad de identificar a los pacientes enfermos. Indica la proporción de pacientes enfermos a los que el test les dio positivo, y se expresa como porcentaje (0-100%) o fracción (de 0 a 1). Una prueba con alta sensibilidad tiene pocos falsos negativos (pacientes enfermos a los que la prueba no detectó). (Al observar la tabla 2 x 2 se nota que para este cálculo se utiliza la columna de enfermos que es el denominador de la sensibilidad).
- **Sensibilidad= Verdaderos Positivos/ Enfermos= VP/ (VP + FN)**
- **Especificidad (E) o Tasa de Verdaderos Negativos (TVN):** La especificidad es la capacidad de un método diagnóstico de reconocer a las personas sanas. Una especificidad alta evita falsos positivos (pacientes sanos a los que el test cataloga como positivos). (Como el denominador de este cálculo son los sanos se utilizan los datos de la segunda columna).
- **Especificidad= Verdaderos Negativos/ Sanos= VN/ (VN + FP)**

- Poder Predictivo Positivo (PPP) o Valor Predictivo Positivo (VPP): El poder predictivo positivo mide la probabilidad de estar enfermo si la prueba es positiva. Es la proporción de pruebas positivas que son verdaderas o corresponden a pacientes enfermos. Una prueba con alto poder predictivo positivo tiene pocos falsos positivos (resultados positivos equivocados). (En la tabla de 2 x 2 las pruebas con resultado positivo se ubican en la primera fila, la de las pruebas positivas, que es la que se utiliza para este cálculo).
- **PPP= Verdaderos positivos/ Positivos= VP/ (VP + FP)**
- Poder Predictivo Negativo (PPN) o Valor Predictivo Negativo (VPN): El poder predictivo negativo mide la probabilidad que tiene un paciente de estar sano si la prueba le dio negativa. Es la proporción de resultados negativos verdaderos o correspondientes a pacientes sanos. Una prueba con alto poder predictivo negativo tiene pocos falsos negativos (resultados negativos erróneos). (Para este cálculo se utiliza la segunda fila de la tabla de 2 x 2 que corresponde a las pruebas con resultado negativo).
- **PPN: Verdaderos Negativos/ Negativos= VN/ (VN + FN)**

Existe otra medida de capacidad operativa que combina toda la información de la tabla y que es aplicable al caso particular: el cociente de probabilidades o razón de probabilidades.

Razón de probabilidades o coeficiente de verosimilitudes (o "Likelihood Ratio": LR): Expresa la chance de estar enfermo cuando la prueba es positiva (LR positivo) o negativa (LR negativo). Compara cuántas veces la prueba acierta contra las veces que se equivoca:

**LR+: tasa de Verdaderos Positivos/ tasa de Falsos Positivos = tVP/ tFP = S/ 1-E**

**LR -: tasa de Falsos Negativos/ tasa de Verdaderos Negativos = tFN/ tVN = 1-S/ E**

Un LR= 1 significa que las chances de un paciente de estar enfermo o sano son iguales (la prueba no ayuda al diagnóstico para la patología en estudio). Para que un método diagnóstico tenga utilidad se espera que su LR+ sea alto (idealmente > 10) y su LR- sea bajo (< 0,1).

*Tomaremos como ejemplo los resultados obtenidos para la prueba de uriscreen comparada contra el estándar de referencia. Un total de 121 niños fueron incluidos en el estudio, 35 resultaron tener urocultivo positivo (enfermos) y todos dieron positivos para el test rápido, pero también dieron positivo 27 niños que no tenían infección urinaria (urocultivo negativo):*

|                    | Urocultivo positivo | Urocultivo negativo |
|--------------------|---------------------|---------------------|
| Uriscreen positivo | 35                  | 27                  |
| Uriscreen negativo | 0                   | 59                  |
|                    | Enfermos            | Sanos               |

A continuación los resultados de nuestro ejemplo:

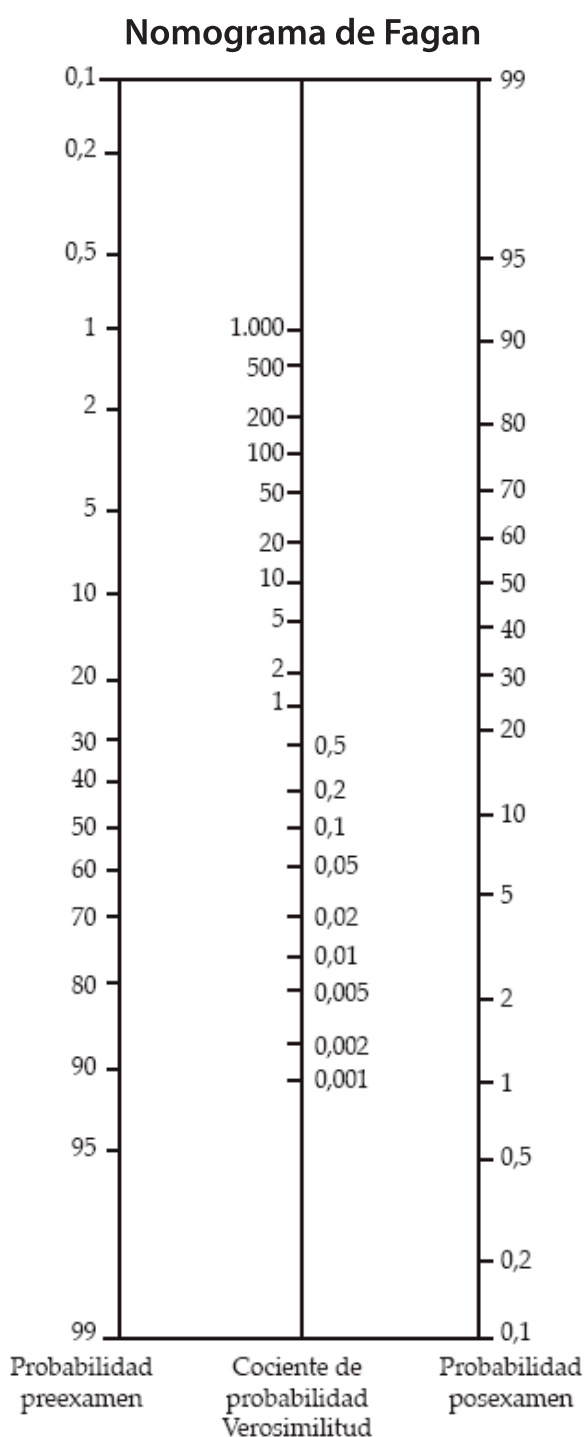
|                                       |               |
|---------------------------------------|---------------|
| <b>Sensibilidad</b>                   | <b>100,0%</b> |
| <b>Especificidad</b>                  | <b>68,6%</b>  |
| <b>Valor predictivo positivo</b>      | <b>56,5%</b>  |
| <b>Valor predictivo negativo</b>      | <b>100,0%</b> |
| <b>Proporción de falsos positivos</b> | <b>31,4%</b>  |
| <b>Proporción de falsos negativos</b> | <b>0,0%</b>   |
| <b>CPP o LR(+)</b>                    | <b>3,19</b>   |
| <b>CPN o LR(-)</b>                    | <b>0,02</b>   |

*Como se podrá interpretar de nuestro ejemplo, el Uriscreen es excelente justamente como método de diagnóstico rápido, ya que tiene una altísima sensibilidad y valor predictivo negativo, lo que permitirá descartar a la gran mayoría de*

## Cambios entre las probabilidades preprueba y postprueba según el cociente de probabilidad.

| Cociente de probabilidad (+) | Cambios en la probabilidad |
|------------------------------|----------------------------|
| >10                          | Grandes y concluyentes     |
| 5-10                         | Moderados                  |
| 2-5                          | Pequeños                   |
| 1-2                          | Insignificantes            |

| Cociente de probabilidad (-) | Cambios en la probabilidad |
|------------------------------|----------------------------|
| 1-0,5                        | Insignificantes            |
| 0,5-0,2                      | Pequeños                   |
| 0,2-0,1                      | Moderados                  |
| <0,1                         | Grandes y concluyentes     |



*los niños sanos, aunque su especificidad es baja, por lo cual a los pacientes con un test positivo los tendré que someter a una prueba confirmatoria (urocultivo).*

Una vez que conocemos los cocientes de probabilidad positivos y negativos de la prueba (recuerde que estos son propios de la capacidad operativa del test y no dependen del resultado) entonces utilizaremos el Nomograma de Fagan, en donde uniremos con una línea recta la probabilidad pretest del paciente (1ra escala vertical a la izquierda) con el cociente de probabilidad del test (2da escala vertical, en el centro del Nomograma). Entonces proyectando la línea hasta la 3ra escala a la derecha del Nomograma nos dará cual será la probabilidad posttest de tener el diagnóstico presuntivo. Para estimar la probabilidad pretest de un paciente, tendremos en cuenta los datos clínicos del paciente o incluso podemos basarnos en la prevalencia de la enfermedad cuando no tenemos ningún dato clínico. También existen programas informáticos para el cálculo de las probabilidades posttest sin la necesidad de utilizar el Nomograma.

El resultado de la prueba aumentará o disminuirá la probabilidad de que el paciente esté enfermo según el LR sea  $> 1$  en el caso de prueba positiva (cuanto mayor el LR mayor la probabilidad) o  $LR < 1$  en el caso de pruebas con resultado negativo (menor la probabilidad de enfermedad a menor LR).

*Como ejemplo podemos tomar una probabilidad pretest del 50% de que un paciente tenga infección urinaria. Si el test de Uriscreeen diera positivo la probabilidad posttest subiría hasta el 80% aproximadamente ( $LR+ 3,19$ ). En cambio para un resultado negativo la probabilidad*

postest bajaría hasta menos del 2 % (LR- 0,02). (Tome una regla y trace una línea desde la probabilidad pretest de 50% hasta la probabilidad postest, pasando en el primer caso por el cociente de probabilidad positivo de 3,19 y luego por el cociente de probabilidad negativo de 0,02).

### Resultados no dicotómicos Curvas ROC

Las curvas ROC o Receiver Operator Characteristics se utilizan para los estudios donde el resultado no es dicotómico (positivo o negativo), sino medidas graduales.

Se grafican los diferentes puntos de corte de un método relacionando la Tasa de verdaderos positivos o sensibilidad (en las ordenadas) y la tasa de falsos positivos o 1 – Especificidad (en las abscisas). Esto permite no sólo comparar puntos de corte de un método determinado sino también comparar dos métodos diferentes.

**Cuanto mayor es la sensibilidad menor es el número de falsos negativos.**

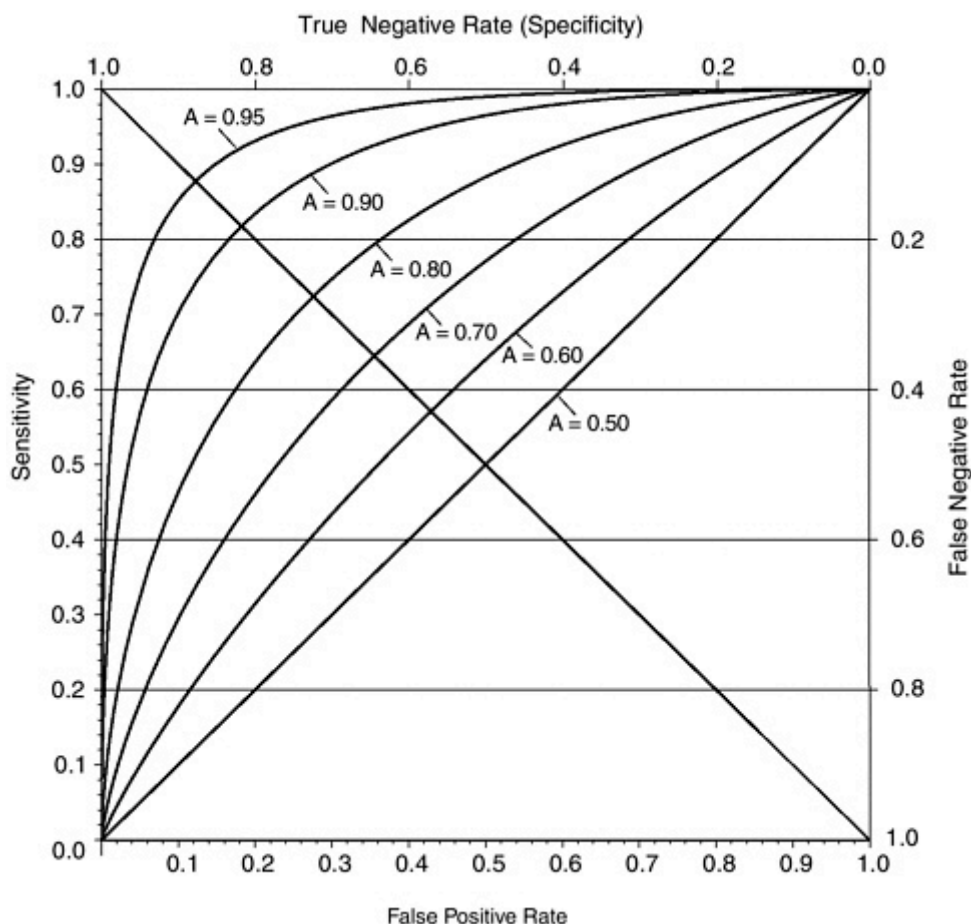
**Cuanto mayor es la especificidad menor es el número de falsos positivos.**

El área total del gráfico es = 1. Una línea diagonal que atravesase el gráfico en dos partes iguales (2 áreas de 0,5), representa el 50% de sensibilidad y 50% de especificidad (lo mismo que tirar una moneda). Por lo tanto, si el área abarcada por la curva ROC o su intervalo de confianza es del 50% (0,5) se puede determinar que el método no es de utilidad para arribar a un diagnóstico, ya que el resultado es azaroso.

El punto óptimo de verdaderos positivos con un mínimo de falsos positivos es el extremo superior izquierdo del gráfico.

| Área bajo la curva | Contribución al diagnóstico |
|--------------------|-----------------------------|
| 0,5 (50%)          | Nula                        |
| 0,5-0,6            | Mala                        |
| 0,6-0,7            | Pobre                       |
| 0,7-0,8            | Aceptable                   |
| 0,8-0,9            | Buena                       |
| 0,9-0,95           | Muy buena                   |
| 0,95-1             | Excelente                   |

### Interpretación de la curva ROC: <sup>7,8</sup>



## Validez externa: ¿Me ayudarán los resultados en la asistencia de mis pacientes?

*(a) ¿Son aplicables los resultados a mi paciente?*

*(b) ¿Está la prueba disponible, es accesible, reproducible y fácilmente interpretable?*

*(c) ¿Los resultados modificarán mi tratamiento?*

*(c) ¿Obtendrán beneficio los pacientes como consecuencia del examen?*

**(a)** En este caso se debe verificar si las poblaciones donde se realizó la prueba y donde se utilizará son similares como para presumir que ésta se comportará de manera similar. Para esto sirve controlar los criterios de inclusión y de exclusión.

**(b)** Aquí se debe evaluar la viabilidad de realizar la prueba en su medio de trabajo, es decir si no es excesivamente costosa, riesgosa e inclusive objetable por los pacientes.

**(c)** Los resultados de las pruebas diagnósticas serán de utilidad especialmente si modifican conductas médicas.

Es decir cuando hay una alta presunción clínica de enfermedad (alta probabilidad pretest) deberá tener un muy alto LR(-) para descartar enfermedad y viceversa. Por lo tanto es fundamental conocer los cocientes de probabilidad de los test diagnósticos antes de decidir realizarlos, en especial en los casos que estos tengan un alto costo económico y/o emocional para los pacientes y el sistema de salud.

**(d)** Para contestar este punto, será necesario medir el impacto de la realización de una prueba diagnóstica como si esta fuera una intervención, es decir medir el impacto de su uso sobre la salud de los pacientes.

## Bibliografía

1. Jaeschke R, et al. Guía del usuario para artículos sobre diagnóstico. JAMA 1994; 271: 389-391, 703-707.
2. Ochoa Sangrador C, González de Dios J, Buñuel Álvarez JC. Evaluación de artículos científicos sobre pruebas diagnósticas. Evid Pediatr. 2007; 3:24.
3. Waisman Y; Elisheva Z; Amir L; Mimouni M. The Validity of the Uriscreeen Test for Early Detection of Urinary Tract Infection in Children. Pediatrics 1999; 104 (4). URL: <http://www.pediatrics.org/cgi/content/full/104/4/e41>
4. Abaira V, Indices de rendimiento de las pruebas diagnósticas. SEMERGEN 2002;28(4):193-94
5. Abaira V. Sesgos en los estudios sobre pruebas diagnósticas. SEMERGEN. 2006;32(1):24-6
6. García GJJ. Significado y empleo de la razón de probabilidades en la práctica clínica. Rev Mex Pediatr. 2000;67(4):188-191
7. Whiting PF, Sterne JA, Westwood ME, Bachmann LM, Harbord R, Egger M, Deeks JJ. Graphical presentation of diagnostic information. BMC Med Res Methodol. 2008 Apr 11;8:20.
8. Florkowski CM. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. Clin Biochem Rev. 2008 Aug;29 Suppl 1:S83-7.