

Estrategias de aprendizaje

## MEDICINA BASADA EN EVIDENCIAS: PRUEBA DIAGNÓSTICA

Salvador Espino y Sosa,\* Myrna Godines Enríquez,\*\* José Antonio Ramírez Calvo \*\*\*

### Resumen

La evaluación de artículos de pruebas diagnósticas, exige el desarrollo de algunas herramientas epidemiológicas sumamente interesantes y útiles para el clínico. La adecuada interpretación de las capacidades de las pruebas conduce a una práctica diaria científica y objetiva. El presente artículo abordará la evaluación de la validez de este tipo de estudios; desarrollaremos en una forma amigable la interpretación de las pruebas y aterrizaremos la factibilidad de su aplicación en una población como la nuestra. Esperamos que esta segunda entrega de la serie de medicina basada en evidencias, sea especialmente útil y amena en su lectura.

**Palabras clave:** Medicina basada en evidencias, diagnóstico, prueba diagnóstica.

### Abstract

The evaluation of diagnostic test articles demands the development of certain epidemiologic tools that are very interesting and useful to the clinician. The right interpretation of the test capacities leads to a scientific and objective daily practice. The paper will explain this kind of articles evaluation validity, also will develop a friendly interpretation of the tests and explain the feasibility of its application in our population. We hope this second issue of the evidence based medicine sequence be specially useful and pleasant in your lecture.

**MeSH words:** Evidence-based medicine, diagnosis, diagnostic test.

*“La medicina es una ciencia de probabilidades y el arte de manejar la incertidumbre”*

Salvador Pita Fernández

### El proceso diagnóstico

La práctica clínica habitualmente inicia con el reto diagnóstico, los procesos del pensamiento que abordan, organizan, sintetizan la información para diagnosticar, utilizan una actividad global del sistema cognitivo con intervención de los mecanismos de memoria, atención, procesos de comprensión, aprendizaje, etc. Este proceso por lo general no es consciente en los médicos experimentados; el abordaje estructurado de diagnósticos diferenciales es natural en el. El proceso puede ser dividido en dos pasos:

En primer lugar, nos formamos un concepto de normalidad y un rango de la misma. Por arriba de un

horizonte arbitrario consideramos la posibilidad de enfermedad. El interrogatorio y la exploración física nos llevarán a identificar anormalidades, factores de riesgo o factores pronósticos. Cuando abordamos a un paciente en cualquier escenario clínico, lo interrogamos y exploramos; en este sentido lo sometemos a la primera batería de pruebas diagnósticas. Los signos y síntomas que recabaremos nos sirven para establecer una probabilidad inicial de ciertas patologías y en este momento empezamos a jugar con la incertidumbre; una pregunta general nos lleva a preguntas más particulares, conduciendo la investigación primero de lo general a lo particular. En este momento, el clínico establece una serie de probabilidades diagnósticas y la probabilidad de cada diagnóstico diferencial es distinta. Probablemente en la mayoría de los casos no lo hagamos consciente y en menor medida objetivo; con poca frecuencia establecemos explícitamente la probabilidad de la enfermedad antes de someter a nuestro paciente a pruebas diagnósticas. Por ejemplo, preguntarnos cual es la probabilidad de que una paciente en urgencias curse con infección de vías urinarias dado que refiere frecuencia urinaria. Seguramente esta probabilidad será muy diferente si la paciente está embarazada o si no lo está. En el caso de tratarse de una mujer embarazada, este síntoma no nos dirá mucho (la probabilidad de infección de vías urinarias en una paciente embarazada no es diferente entre mujeres que refieren frecuencia y las que

\* Dirección de investigación

\*\* Dirección de investigación

\*\*\* Departamento de Especialidades Médicas

Instituto Nacional de Perinatología Isidro Espinosa de los Reyes.

Correspondencia:

Tel. (55) 55209900 ext. 375

salvadorespino@gmail.com

no lo hacen), pero si no lo está, el síntoma adquiere un peso diferente y obliga a sospechar infección de vías urinarias. En este mismo sentido, la probabilidad de la infección es diferente si los síntomas son diferentes, la presencia de disuria se relacionará con una mayor probabilidad de infección, que la presencia de frecuencia. Pero ¿qué tan diferente? Eso generalmente no lo definimos estadísticamente, ya que solo aceptamos intuitivamente que existe una diferencia.

En segundo lugar, una vez que establecemos un espectro de posibilidades para nuestro diagnóstico, optamos por implementar pruebas que nos orienten hacia alguno en particular; es decir, pruebas orientadas a aumentar o a disminuir la probabilidad de cada opción de salud o de enfermedad. En el caso de la paciente en urgencias con síntomas urinarios, optamos por determinar una tira reactiva en orina, cuyos resultados elevan la probabilidad diagnóstica o la disminuyen. En este sentido, estamos lejos de estar en medio de un juego de todo o nada, en medicina generalmente jugamos con probabilidades y nuestros esfuerzos están orientados a reducir la incertidumbre. Cuando conocemos las propiedades de cada paso en el proceso diagnóstico, podemos hacer más objetiva nuestra práctica clínica y sustentar de forma más real nuestra toma de decisiones. Como veremos posteriormente, el proceso diagnóstico se reduce a:

1. Establecer una probabilidad de la enfermedad previa a la prueba
2. Determinar el grado con el que el resultado de la prueba modifica esta probabilidad basal
3. Interpretar la probabilidad posterior a la prueba para tomar una decisión clínica

**Tabla 1.** Abordaje de artículos de pruebas diagnósticas

Validez de los resultados
<ul style="list-style-type: none"> <li>• ¿El artículo, realmente investiga la incertidumbre diagnóstica?</li> <li>• ¿La prueba diagnóstica fue comparada con un estándar de oro de forma enmascarada?</li> <li>• ¿El resultado de la prueba diagnóstica motivó la realización o no del estándar de oro?</li> </ul>
Capacidades de la prueba
<ul style="list-style-type: none"> <li>• ¿Qué capacidad tiene la prueba diagnóstica para detectar o descartar la enfermedad?</li> </ul>
Aplicabilidad de la prueba
<ul style="list-style-type: none"> <li>• ¿Es factible la reproducibilidad y la interpretabilidad de la prueba en mi medio?</li> <li>• ¿Los resultados de la prueba son aplicables a mi práctica clínica?</li> <li>• ¿Los resultados motivarán un cambio en la toma de decisiones clínicas?</li> <li>• ¿La prueba mejorará las condiciones de mi población?</li> </ul>

Un reto al que se enfrenta el clínico, es el de establecer una secuencia eficaz de pruebas con un sustento lógico para no hacer pruebas, innecesarias, costosas o peligrosas. La intención en este sentido es escoger la o las pruebas cuyo resultado conduzca a una modificación sustancial de la probabilidad de la enfermedad sobre una base ética, tomando en cuenta la autonomía del paciente y sus preferencias.

Los artículos que nos hablan de pruebas diagnósticas deben ser cuestionados (como se comentó en el primer artículo de esta serie) en términos de su validez, las capacidades de la prueba y de su aplicabilidad al paciente o población que tratamos; para esto nos basamos en un escrutinio dirigido (Tabla 1).

### ***Determinación de la validez de un artículo de prueba diagnóstica***

#### **¿El artículo, realmente investiga la incertidumbre diagnóstica?**

Esta pregunta lleva a dos reflexiones:

En primer lugar, las enfermedades usualmente se presentan con un espectro de gravedad; cuando tenemos un paciente que cursa con una forma avanzada de la enfermedad, no necesitamos una prueba adicional a nuestra anamnesis para diagnosticar la entidad. Las formas avanzadas de las enfermedades se acompañan de una importante serie de manifestaciones clínicas (a veces patognomónicas) cuya identificación al interrogatorio y a la exploración, permite el diagnóstico preciso y la toma de decisiones terapéuticas inmediatas. En este grupo de pacientes, la utilidad clínica de pruebas diagnósticas es prácticamente nula. Dicho en otras palabras, en los pacientes que cursan con formas avanzadas de la enfermedad, podemos usar nuestro juicio clínico para establecer una alta probabilidad de la entidad diagnóstica sospechada, probabilidad suficiente para dar el paso al tratamiento y optar por pruebas pronósticas más que diagnósticas.

En segundo término, cuando una prueba diagnóstica es evaluada en una población gravemente enferma, de manera habitual resulta muy buena para identificar enfermos y tiene una alta probabilidad de resultar positiva (sensibilidad; se revisará más adelante), es decir el espectro de la enfermedad modifica la sensibilidad de la prueba. De la misma manera, las pruebas que se evalúan en población poco enferma resultan menos sensibles. Así, si el artículo evalúa la prueba diagnóstica en una población gravemente enferma, estará reportando una sensibilidad más alta de la que tendría la misma prueba si se aplica en una población

con formas menos graves de la enfermedad. La verdadera utilidad de la prueba estribará en el hecho de si sirve en la práctica clínica cotidiana y no solamente en un grupo de pacientes especialmente enfermos. La prueba diagnóstica será útil cuando permita identificar de forma temprana la presencia de la enfermedad por lo que el estudio debe contemplar una muestra amplia en el espectro de la enfermedad; es decir, que distinga entre estados que puedan llegar a ser confusos clínicamente.

El ejemplo ya clásico, es el de la historia de la interpretación del antígeno carcinoembrionario en pacientes con cáncer colorectal. Cuando se evaluaron los resultados de estas concentraciones en 36 pacientes con cáncer colorectal avanzado, el antígeno carcinoembrionario resultó elevado en 35 de ellos lo que supuso a la comunidad médica que su determinación resultaría de utilidad para el diagnóstico. Al hacer estudios en población con estadios tempranos, las concentraciones de antígeno carcinoembrionario no mostraron diferencias significativas con la población normal, lo que condujo al abandono de esta determinación como prueba diagnóstica (aunque ahora se usa en el seguimiento de estos pacientes).

La determinación de estereasa leucocitaria y nitritos bacterianos en tira reactiva, muestra una sensibilidad muy diferente, cuando se compara la determinación en pacientes con alta sospecha (92%, IC95% 82 a 98%) y baja sospecha de infección de vías urinarias (52%, IC95% d3 a 79%), por lo que no podemos interpretar de la misma forma el resultado, cuando realizamos la determinación en poblaciones con diferente espectro de la enfermedad).

### **¿La prueba diagnóstica fue comparada con un estándar de oro de forma enmascarada?**

Un estándar de oro es la prueba que ha demostrado mayor sensibilidad y especificidad para el diagnóstico de una enfermedad y por lo tanto es la más aceptada como criterio de verdad. Es la prueba considerada como definitiva. Hipotéticamente, un estándar de oro ideal tiene una sensibilidad del 100% (resulta positiva en todos los casos de enfermedad) y una especificidad del 100% (resulta negativa en todos los casos en ausencia de enfermedad). Sin embargo, en la práctica clínica es difícil encontrar estándares de oro con estas características. Así, el estándar de oro real es la prueba más sensible y específica que sea accesible y factible de utilizar.

El artículo que evalúa una prueba diagnóstica, debe comparar el resultado en todos y cada uno de los pacientes, con el estándar de oro establecido para la enfermedad en cuestión (el lector del artículo debe eva-

luar el estándar de oro establecido y determinar si es la mejor opción para comparar la prueba). Los valores de sensibilidad y especificidad estarán en función de los resultados comparados con el estándar de oro. Así, los resultados de la prueba serán contrastados con la verdad (o el grado de verdad establecido por el estándar de oro).

Idealmente, la prueba a evaluar debe de ser independiente del estándar de oro. Cuando el estándar de oro es conformado por una serie de pruebas diagnósticas que incluyen la prueba a evaluar, el poder de la prueba diagnóstica puede ser sobrevalorado porque su resultado influye en el resultado del estándar de oro.

Una vez que el lector ha aceptado el estándar de oro como una buena prueba de comparación y ha asegurado la independencia con la prueba diagnóstica, debe cerciorarse que la interpretación de la prueba y del estándar hayan sido enmascaradas; es decir, que la persona que interpretó el resultado de la prueba diagnóstica, no conozca el resultado del estándar en el mismo paciente y viceversa. La falta de cegamiento puede generar un sesgo de interpretación. Cuando un clínico sabe el resultado del estándar de oro, tiende a interpretar la prueba diagnóstica en el mismo sentido, lo que resulta en un falso aumento de la sensibilidad y de la especificidad. El conocimiento del resultado de una prueba puede influir en la interpretación de otra.

Otro sesgo que debemos de tomar en cuenta en el diagnóstico, es el que ocurre cuando la frecuencia o número de pruebas diagnósticas es diferente entre la población que resultó enferma y la población catalogada como sana por el estándar de oro; el hecho de realizar más pruebas en la población enferma conducirá a una mayor probabilidad de que alguna resulte positiva.

### **¿El resultado de la prueba diagnóstica motivó la realización del estándar de oro?**

Existen ciertas situaciones en las que se decide realizar el estándar de oro solo cuando el resultado de una prueba es positivo; un ejemplo de ello, es cuando el estándar de oro es una prueba costosa o no inocua y el clínico prefiere reservarla para un grupo de pacientes previamente seleccionados por la prueba a evaluar. Cuando esto sucede ocurre el llamado sesgo de verificación. En estos casos puede suceder que el estándar de oro se realice solo en pacientes con la prueba positiva y que no haya forma de evaluar su poder, para determinar ausencia de enfermedad.

### **Capacidades de la prueba**

La correcta interpretación de una prueba diagnóstica, exige el conocimiento de sus capacidades para de-

tectar o descartar la enfermedad. Esta capacidad está en función de su correlación con el estándar de oro. Una prueba ideal resulta positiva en todos los casos de enfermedad y negativa en todos los casos de ausencia de enfermedad, pero esto resulta utópico en la práctica clínica diaria. Para evaluar una prueba, la comparamos con el estándar de oro (Tabla 2). A partir de la tabla de contingencia podemos calcular las capacidades de la prueba.

- La sensibilidad es la capacidad que tiene la prueba para detectar a los pacientes enfermos; es decir, es la relación de verdaderos positivos entre el total de enfermos, la calcularíamos dividiendo:  $a / a + c$
- La especificidad es la capacidad que tiene la prueba para detectar a los pacientes sanos, es decir, es la relación de verdaderos negativos entre el total de sanos, la calcularíamos dividiendo:  $d / b + d$
- El valor predictivo positivo (VPP) es la probabilidad de que un paciente con la prueba positiva esté realmente enfermo; es decir, es la relación de verdaderos positivos entre el total de pacientes con la prueba positiva, la calculamos dividiendo:  $a / a + b$
- El valor predictivo negativo (VPN) es la probabilidad de que un paciente con la prueba negativa esté realmente sano, es decir, es la relación de verdaderos negativos entre el total de pacientes con la prueba negativa, la calculamos dividiendo:  $d / c + d$

La sensibilidad y la especificidad son capacidades estables de las pruebas diagnósticas; es decir no se modifican con cambios en la prevalencia de la enfermedad, aunque como ya se comentó, estas capacidades sí son modificadas por diferentes puntos en el espectro de gravedad de la enfermedad. Los valores predictivos sí son modificados por prevalencias dife-

rentes, por lo que una prueba diagnóstica mostrará un mayor VPP, cuando se aplica a una población con una prevalencia alta de la enfermedad, que cuando se aplica a una población con una prevalencia baja. En este sentido, el lector debe considerar valores predictivos diferentes, si la enfermedad tiene una prevalencia diferente en su práctica clínica que la reportada por el estudio en cuestión.

Es importante señalar que la interpretación de las capacidades de la prueba, tiene una aplicación diferente en la práctica clínica. Cuando escogemos una prueba diagnóstica para ser aplicada, nos interesan la sensibilidad y la especificidad, ya que para una situación en particular escogeremos una prueba más sensible o más específica, según se requiera. Cuando nos interesa tamizar a la población para una enfermedad, optamos por una prueba muy sensible, porque la intención primordial es detectar a todos los enfermos, aunque consideremos a muchos pacientes realmente sanos como enfermos. Posteriormente seleccionaremos una prueba muy específica para excluir a los sanos que con la prueba de tamizaje fueron considerados enfermos falsamente. Por ejemplo, durante el embarazo aplicamos una prueba de tamiz de glucosa; con ella detectamos a las pacientes que potencialmente tienen diabetes gestacional, pero no hacemos el diagnóstico hasta que una curva de tolerancia oral a la glucosa lo confirma (aplicamos una prueba muy sensible y posteriormente una prueba muy específica). Un caso distinto es el ambiente clínico, ya que cuando estamos ante el paciente y debemos decidir si está enfermo o no, interpretamos los valores predictivos y no la sensibilidad y especificidad, por que como es lógico, realmente no sabemos si el paciente está enfermo o sano; solo sabemos el resultado de la prueba y queremos saber si el sujeto está enfermo o sano, dado que la prueba es positiva o negativa.

Como lo mencionamos, los valores predictivos son modificados por la prevalencia de la enfermedad y esto dificulta su interpretación, porque es difícil que la prevalencia en nuestra población sea igual a la prevalencia de la población estudiada y aun más, la probabilidad de que nuestro paciente tenga la enfermedad es una probabilidad particular en cada caso y no encontraremos estudios diseñados para poblaciones con las características específicas del paciente que tenemos en frente. Una excelente alternativa que tenemos para interpretar las pruebas con las particularidades de nuestros pacientes, es el uso de razones de verosimilitud.

Las razones de verosimilitud (RV) son calculadas a partir de las capacidades estables de la prueba: sensibilidad y especificidad, por lo que no son influidas

**Tabla 2.** Tabla de contingencia de pruebas diagnósticas

		Estándar de oro		
		Positivo	Negativo	
Prueba nueva	Positiva	Verdaderos positivos (a)	Falsos positivos (b)	Total de pacientes con prueba positiva
	Negativa	Falsos negativos (c)	Verdaderos negativos (d)	Total de pacientes con prueba negativa
		Total de enfermos	Total de sanos	

por diferentes prevalencias de la enfermedad. Cuando un paciente presenta la prueba positiva, es de gran utilidad saber cuál es la probabilidad de que este resultado se presente en la población de enfermos, en relación con la probabilidad de que este resultado se presente en la población de sanos. Esta razón de probabilidades se obtiene dividiendo la sensibilidad entre el complemento de la especificidad. Si tomamos en cuenta la tabla de contingencia, la calcularíamos de la siguiente forma  $(a / a + c) / (b / b + d)$ . El mismo interés tendríamos cuando una prueba resulta negativa; el cálculo lo realizaríamos dividiendo el complemento de la sensibilidad entre la especificidad en nuestra tabla de contingencia:  $(c / a + c) / (d / b + d)$ . La correcta interpretación de las RV exige que establezcamos una probabilidad previa a la prueba cercana a la realidad y que esta probabilidad la modifiquemos por la razón de verosimilitud positiva o negativa (según haya resultado la prueba), para obtener una probabilidad de la enfermedad posterior a la prueba (Figura 1).

El cálculo matemático de la modificación de la probabilidad de enfermedad se realiza factorizando los momios de enfermedad por la RV, lo que hace complicado su uso en la práctica diaria, sobre todo en un servicio como el de urgencias, donde el tiempo es un factor determinante. Una buena alternativa es utilizar el nomograma de Fagan (Figura 2)

El nomograma de Fagan facilita la conversión de la probabilidad previa a la prueba, en la posterior a la prueba: En la primera columna identificaríamos la probabilidad de que nuestro paciente tenga la enfermedad antes de realizar la prueba diagnóstica; contiene una secuencia de porcentajes que va del 0.1 al 99%. Esta probabilidad es establecida con los hallazgos del interrogatorio y de la exploración física que hemos realizado o simplemente, con la prevalencia de la enfermedad en la población general. Una vez que hemos determinado el valor basal lo identificamos en el nomograma de Fagan, en la primera columna.

Si realizamos la prueba diagnóstica, el resultado puede ser una RV positiva o negativa (dependiendo del resultado positivo o negativo de la prueba), el valor de la RV es identificado en el nomograma en la segunda columna y trazamos una línea recta desde la probabilidad previa a la prueba, hasta el valor de la RV; esta recta la prolongamos hasta la columna de la probabilidad posterior a la prueba e identificamos en el punto de intersección de ambas columnas, la probabilidad posprueba.

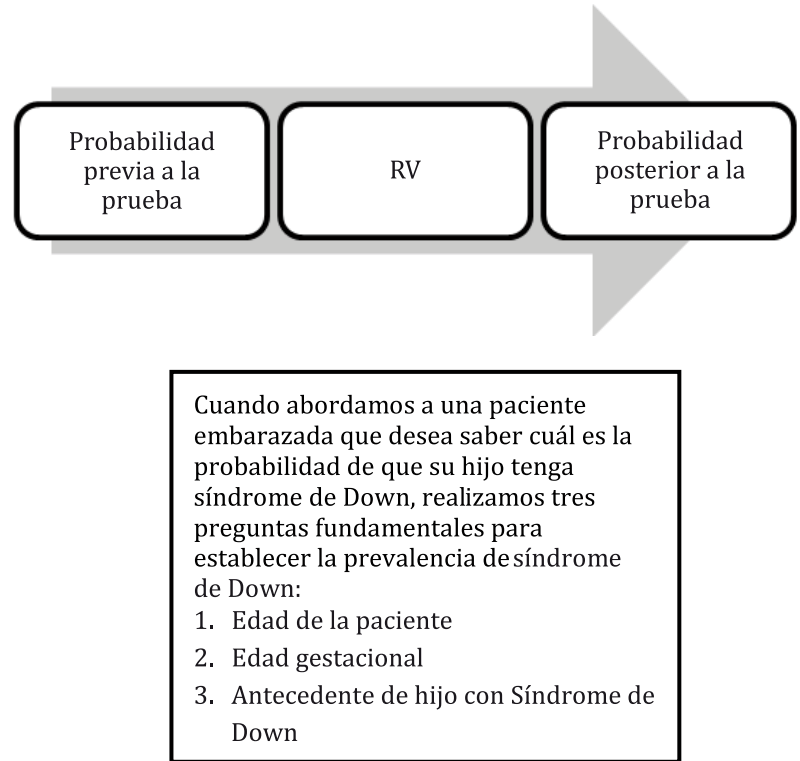
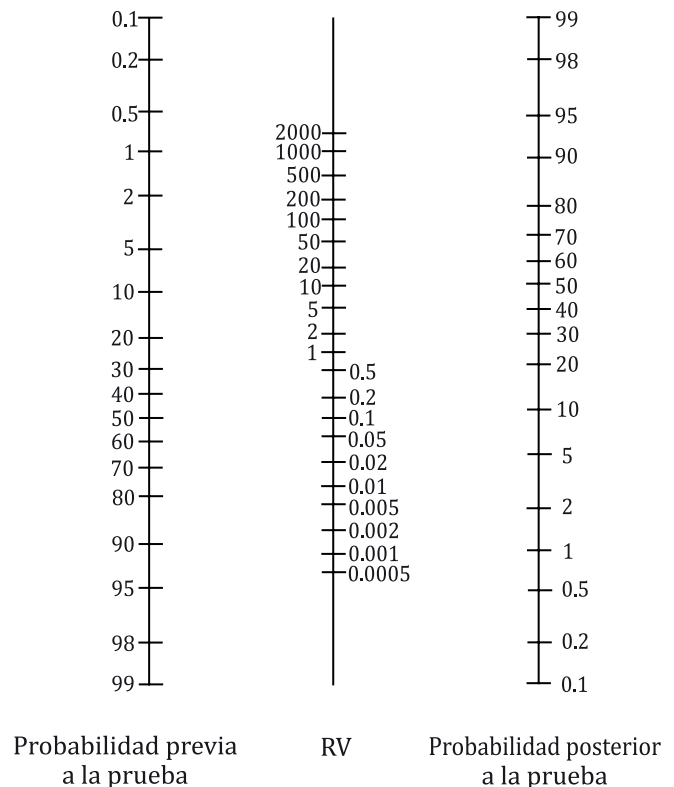
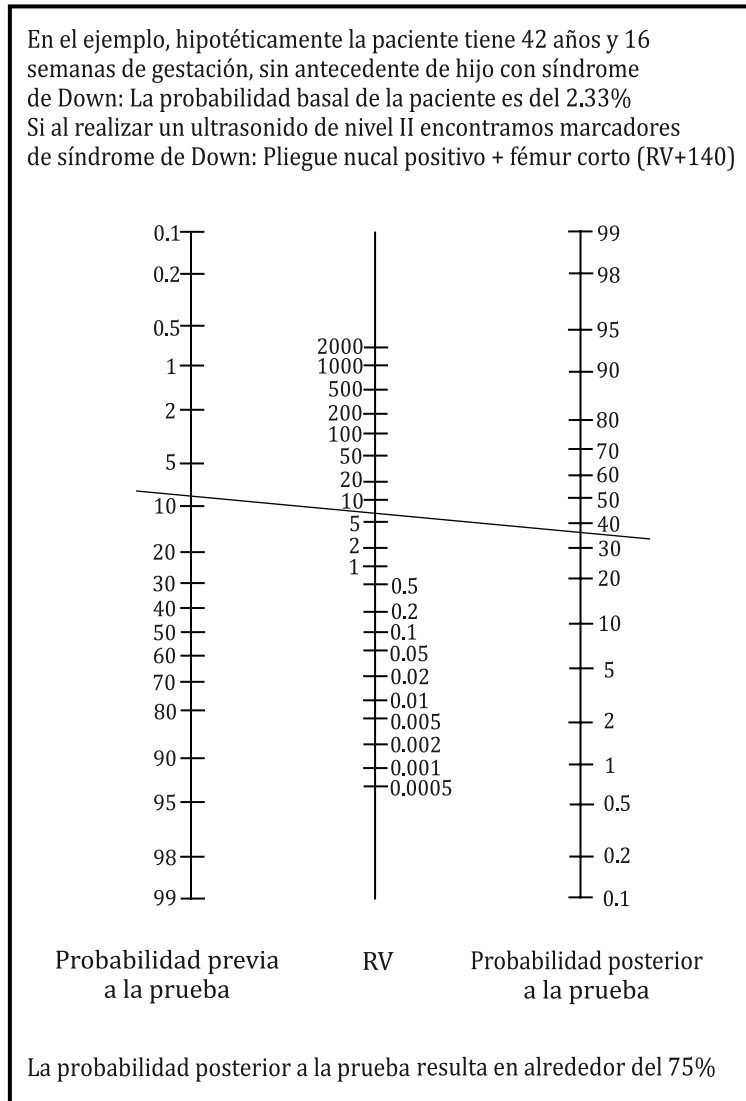


Figura 1. Secuencia en la interpretación de RV (razones de verosimilitud)







**Figura 2.** Nomograma de Fagan.

### ¿Qué capacidad tiene la prueba diagnóstica para detectar o descartar la enfermedad?

Así, podemos deducir que las pruebas que tienen razones de verosimilitud más extremas, son mejores para elevar la sospecha de enfermedad o disminuirla. En este sentido, podemos interpretar la influencia de las RV en función de la siguiente guía:

- Razones de verosimilitud mayores a 10 o menores a 0.1 ocasionan cambios importantes y generalmente concluyentes entre las probabilidades previas y posteriores a la prueba.
- Razones de verosimilitud de 5 a 10 o de 0.2 a 0.1 ocasionan cambios entre las probabilidades previas y posteriores a la prueba.

- Razones de verosimilitud de 2 a 5 o de 0.5 a 0.2 ocasionan cambios pequeños entre las probabilidades previas y posteriores a la prueba.
- Razones de verosimilitud de 1 a 2 o de 1 a 0.5 ocasionan cambios que clínicamente rara vez son importantes entre las probabilidades previas y posteriores a la prueba.

Un punto importante a resaltar, es que el uso de razones de verosimilitud exige el conocimiento o un cálculo realista de la probabilidad basal y, esta probabilidad es modificada por los hallazgos al interrogatorio y a la exploración física. Una excelente referencia para conocer el poder de los hallazgos clínicos se halla publicada en JAMA, en un artículo del Dr. David Sackett y se pueden encontrar en la literatura médica, revisiones sistemáticas sobre el valor diagnóstico del interrogatorio y la exploración física.

### *Aplicabilidad de la prueba diagnóstica*

#### **¿El equipo clínico, es capaz de implementar la prueba diagnóstica en mi medio?**

Una condición esencial para aplicar la prueba diagnóstica a nuestros pacientes es que la prueba pueda ser realizada e interpretada de forma confiable. Cuando una prueba requiere de un experto con un alto nivel de entrenamiento para su aplicación, la factibilidad de su implementación se ve afectada, ya que los clínicos en el medio habitual, no obtendrán los mismos resultados que los expertos que implementaron la prueba en el artículo analizado.

En el mismo sentido, si una prueba tiene una mala reproducibilidad, ocasionará resultados poco válidos y la curva de aprendizaje del operador será muy grande; el intérprete requerirá de mucho tiempo para lograr resultados válidos, ya sea por la dificultad en la técnica o en la interpretación. La baja reproducibilidad de las pruebas, resulta en interpretaciones contrarias entre observadores restándole aplicabilidad a la prueba. Por el contrario, cuando una prueba es altamente reproducible y objetiva, usualmente es fácil de interpretar, no se requiere de observadores altamente entrenados y la curva de aprendizaje para desarrollar la prueba es corta. Cuando una prueba reúne estas características será fácilmente aplicable a nuestro medio.

#### **¿El resultado es aplicable a la práctica diaria?**

Al llegar a este punto en el análisis de un artículo de prueba diagnóstica, debemos hacer dos consideraciones muy interesantes: evaluar el espectro de la enfermedad en pacientes que fueron sometidos a la

prueba diagnóstica y la prevalencia de la misma en la población estudiada.

Los resultados encontrados para la sensibilidad y especificidad pueden ser llamados índices nosológicos, ya que han sido extraídos de pacientes en los que el diagnóstico era conocido. Lo que en realidad necesita el clínico son índices que puedan ser aplicables a pacientes en los que no se ha establecido la presencia o ausencia de enfermedad y que la aplicación pueda ser implementada en el escenario real de trabajo.

Las pruebas diagnósticas son más capaces de detectar enfermedad cuando ésta es más grave, es decir, la sensibilidad de las pruebas diagnósticas es mayor cuando se busca enfermedad en pacientes severamente enfermos, de lo que se deduce que el espectro de la enfermedad, modifica la sensibilidad de la prueba diagnóstica. También es importante mencionar qué variaciones en los estados clínicos, patológicos y comórbidos modifican los valores de sensibilidad y especificidad en diferentes puntos del espectro de la enfermedad y en los estados complementarios de ausencia de enfermedad.

Cuando la muestra se compone en su mayoría por pacientes con formas graves de la enfermedad, la sensibilidad de la prueba se incrementará y por lo tanto, la razón de verosimilitud positiva se alejará de la unidad. Estos valores podrán ser válidos para hospitales de concentración, servicios de atención especializada o terapias intensivas (donde se concentran pacientes graves), pero probablemente no lo serán para clínicas de primer contacto y primer nivel de atención, ya que el grupo de pacientes que suele acudir a estas clínicas, presentan con mayor frecuencia formas menos graves de la enfermedad. Cuando los pacientes sanos cursan con condiciones que la prueba puede interpretar como enfermedad, se incrementa la tasa de falsos positivos disminuyendo la especificidad; esto afecta la razón de verosimilitud negativa acercándola a la unidad y por lo tanto la capacidad de la prueba para detectar sanos.

La interpretación de los valores predictivos de las pruebas diagnósticas exige tomar en cuenta la prevalencia de la enfermedad en la población estudiada en el artículo y compararla con la prevalencia de nuestra población. Recordemos que cuando la prevalencia de la enfermedad es alta, el valor predictivo positivo de la prueba también lo será y el valor predictivo negativo de la prueba será bajo, a diferencia; cuando la prevalencia es baja, el valor predictivo negativo de la prueba será alto y el valor predictivo positivo será bajo.

Si realizáramos un perfil biofísico fetal a la población abierta para sospechar asfixia (su prevalencia es del 0.4%), obtendríamos un valor predictivo positivo del

8% (es decir: ¡solo 8 de cada 100 pacientes identificados como con probable asfixia estarían realmente asfixiados!). Por otra parte, si realizamos el perfil biofísico fetal en una población de alto riesgo, en donde las pacientes fueron identificadas previamente por una prueba sin estrés no reactiva y donde la prevalencia de asfixia es del 30 al 40%, obtendremos un valor predictivo positivo del 92 al 93%. De lo anterior se desprende la importancia de identificar la prevalencia de la enfermedad en la población estudiada y compararla con la prevalencia de la enfermedad en nuestra práctica clínica diaria.

Cuando el paciente a quien pretendemos aplicar la prueba diagnóstica cumple con los criterios de selección de la muestra estudiada, podemos confiar en que las capacidades de la prueba serán aplicables al paciente.

### **¿El resultado de la prueba realmente cambiará el tratamiento en el paciente?**

La finalidad de las pruebas diagnósticas es reducir la incertidumbre para establecer, verificar o terminar una ruta terapéutica. Este proceso tiene dos componentes: al primer componente le exigimos cierto grado de confiabilidad para poder creer en los resultados de la prueba, esto lo verificamos con la validez de las capacidades reportadas de la prueba y con la interpretación directa de las mismas. El segundo componente es interpretado como dos umbrales de decisiones clínicas: Tenemos un umbral diagnóstico. Por debajo de este decidimos que la enfermedad está descartada y por lo tanto, no continuamos haciendo pruebas diagnósticas o suspendemos un tratamiento establecido, y tenemos un umbral terapéutico; por arriba de este, decidimos que tenemos pruebas suficientes de la existencia de la enfermedad, como para justificar el inicio del tratamiento o su consecución. Las pruebas diagnósticas solo tienen justificación entre estos dos umbrales. Las capacidades de la prueba que dictan la conducta son los valores predictivos, ya que definen la probabilidad de la existencia de la enfermedad dado el resultado de la prueba diagnóstica, las razones de verosimilitud también permiten modificar la prevalencia de la enfermedad dado el resultado.

Las probabilidades de la prueba que definen los umbrales mencionados, deben ser definidas a priori y están sujetas a consideraciones particulares: Cuando el tratamiento de la enfermedad tiene riesgos para la salud, el clínico opta por escoger valores predictivos positivos altos para iniciar el tratamiento o una probabilidad posprueba muy alta y reservar el tratamiento altamente riesgoso solo para los pacientes con una muy alta sospecha de la patología. Cuando se trata de una enfermedad con un muy mal pronóstico, si no se

trata tempranamente, el clínico trata pacientes con probabilidades menores que el anterior ejemplo; es decir, busca pruebas muy sensibles aunque disminuya la especificidad.

En este sentido podemos considerar dos situaciones: Primero: Cuando el clínico tiene una alta sospecha de la enfermedad y de la ausencia de esta, no está justificado realizar pruebas solo para confirmar la sospecha si el resultado contrario a lo supuesto no va a hacer cambiar la conducta. Segundo: las pruebas diagnósticas con razones de verosimilitud cercanas a uno no van a producir un cambio significativo en la conducta y no van a contradecir la suposición clínica de la presencia o ausencia de enfermedad, en este sentido no tiene caso pedir las.

### **¿La implementación de la prueba beneficiará realmente a mi paciente?**

En este punto se ponen a consideración dos reflexiones:

En el balance riesgo beneficio, mi paciente realmente obtiene un beneficio al realizarse la prueba, o corre más riesgo con la implementación de la misma. En este rubro, no solo nos referimos a los riesgos de la implementación de la prueba diagnóstica por razones de su técnica o su interacción con nuestro paciente, sino también a las consecuencias del tratamiento motivado por el resultado de la prueba. Ya hemos expuesto el caso de la prueba sin estrés aplicado a la población general, donde un resultado positivo para la enfermedad se asociará muy pobremente a la presencia real de la misma, si el clínico actúa en con-

secuencia solamente del resultado de la prueba sin estrés, estará sobre diagnosticando asfixia perinatal y la interrupción del embarazo de forma temprana en fetos no enfermos tendrá complicaciones innecesarias muy importantes. En este sentido será mejor no realizar la prueba diagnóstica que si hacerlo.

La implementación de la prueba debe de ser opción del paciente y no una prerrogativa del médico, en este sentido, tomar en cuenta las preferencias del paciente se torna fundamental, así como su economía. Así, la mejor prueba diagnóstica no es la más sensible y específica sino aquella que además, cumpla con las expectativas del paciente y sea aceptada por él.

### **Referencias bibliográficas**

1. Thomson DMP, Krupey J, Freedman SO, Gold P. The radioimmunoassay of circulating carcinoembryonic antigen of the human digestive system. Proc Natl Acad Sci U S A. 1969; 64:161-167.
2. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Ann Intern Med. 1992 Jul 15;117(2):135-40.
3. Sackett DL, Rennie D. The science and art of the clinical examination. JAMA. 1992;267:2650-2652.
4. Wulff HR. Rational diagnosis and treatment. 2nd ed. Oxford: Blackwell Scientific, 1981.
5. Feinstein AR. Misguided efforts and future challenges for research on "diagnostic tests". J Epidemiol Community Health 2002;56:330-332
6. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Saunders, 1985.