

Cómo interpretar las pruebas diagnósticas

How to interpret diagnostic tests

Ignacio Pérez^a, Iara Yamila Taito-Vicenti^b, Catalina Gracia González-Xuriguera^a,
Cristhian Carvajal^a, Juan Víctor Ariel Franco^b, Cristóbal Loézar^{c,d,*}

^a Escuela de Medicina, Universidad de Valparaíso, Viña del Mar, Chile

^b Instituto Universitario Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

^c Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Viña del Mar, Chile

^d Centro Asociado Universidad de Valparaíso, Cochrane Chile, Viña del Mar, Chile

*Autor corresponsal cristobal.loezar@uv.cl

Citación Pérez I, Taito-Vicenti IY, González-Xuriguera CG, Carvajal C, Franco JVA, Loézar C. How to interpret diagnostic tests. *Medwave* 2021;21(7):e8432

Doi 10.5867/medwave.2021.07.8432

Fecha de envío 28/11/2020

Fecha de aceptación 26/5/2021

Fecha de publicación 4/8/2021

Origen No solicitado

Tipo de revisión Con revisión por pares externa, por dos árbitros a doble ciego

Palabras clave diagnostic test, accuracy, impact, sensitivity, specificity

Resumen

En el ámbito de la salud, los profesionales deben tomar decisiones en un marco de incertidumbre. Al realizar un diagnóstico, se categorizan los signos y síntomas, sumados a los hallazgos de exámenes complementarios de una condición clínica particular, implicando la definición de un tratamiento y pronóstico específico. Durante el proceso diagnóstico se utilizan herramientas de la anamnesis, examen físico y exámenes complementarios para apoyar dicha categorización. Estas herramientas, conocidas como pruebas diagnósticas, permiten estimar la probabilidad de la presencia o ausencia de la condición médica sospechada. La utilidad de las pruebas diagnósticas varía para cada condición clínica y se evalúan mediante estudios de exactitud (sensibilidad y especificidad) e impacto diagnóstico (repercusión sobre los desenlaces de salud). En este artículo, se abordan los conceptos teóricos y prácticos generales sobre las pruebas diagnósticas en seres humanos considerando sus antecedentes históricos, su relación con las teorías sobre probabilidades y su utilidad práctica con ejemplos ilustrativos.

Abstract

Healthcare professionals make decisions in a context of uncertainty. When making a diagnosis, relevant patient characteristics are categorized to fit a particular condition that explains what the patient is experiencing. During the diagnostic process, tools such as the medical interview, physical examination, and other complementary tests support this categorization. These tools, known as diagnostic

tests, allow professionals to estimate the probability of the presence or absence of the suspected medical condition. The usefulness of diagnostic tests varies for each clinical condition, and studies of accuracy (sensitivity and specificity) and diagnostic impact (impact on health outcomes) are used to evaluate them. In this article, the general theoretical and practical concepts about diagnostic tests in human beings are addressed, considering their historical background, their relationship with probability theories, and their practical utility with illustrative examples.

Ideas clave

- Al decidir en marcos de incertidumbre, la mayoría de las veces los profesionales no disponen de una certeza absoluta sobre la condición diagnosticada a un paciente.
- Las pruebas diagnósticas apoyan el proceso diagnóstico en la categorización de las vivencias de un paciente en una condición médica particular que implica una patogénesis, tratamiento y pronóstico específicos.
- Existen distintos tipos de pruebas diagnósticas que pueden ser desde preguntas en la anamnesis, signos al examen físico hasta exámenes complementarios (laboratorio, imágenes u otros procedimientos). Estas son evaluadas mediante estudios de exactitud e impacto.
- Este artículo ofrece un acercamiento a las revisiones disponibles en las principales bases de datos y textos de consulta especializados, referidas a pruebas y exactitud diagnóstica en seres humanos en un lenguaje amigable, orientado a la formación de estudiantes de pre y posgrado.

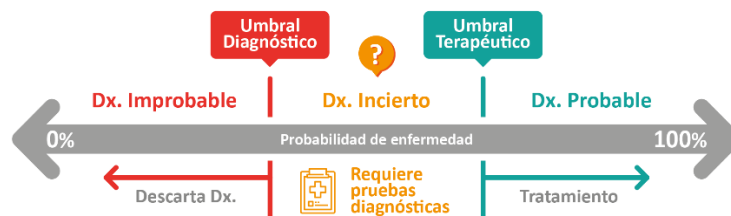
Introducción

En el ámbito de la salud, los profesionales deben tomar decisiones en un contexto de incertidumbre. Al realizar un diagnóstico, los clínicos categorizan las vivencias de un paciente en una condición particular que implica una patogénesis, tratamiento y pronóstico específicos¹. Sin embargo, la mayoría del tiempo no existe una certeza absoluta si es que un paciente tiene realmente la condición que se le ha diagnosticado².

Hace más de un siglo, el diagnóstico se basaba principalmente en la anamnesis y el examen físico. Según Erick Cobo y colaboradores, el monje inglés Thomas Bayes llegó a la conclusión de que la existencia de Dios sólo puede demostrarse si previamente se cree en Dios. Por lo tanto, lo que determina la probabilidad de que exista Dios depende de ser creyente o no³. Este razonamiento aplicado al diagnóstico médico plantea que la probabilidad de un evento posterior a la aplicación de una prueba, depende de la probabilidad que el evento tenía previa a la misma y las características propias de la prueba aplicada⁴. La adjudicación de la probabilidad previa a la aplicación de la prueba es un proceso en el cual el profesional de salud utiliza su conocimiento, experiencia y juicio clínico⁵.

A su vez existen otras aproximaciones diagnósticas como la heurística, definida por Pérez⁶ como “mecanismos psicológicos basados en la actuación humana frente a la resolución de problemas, por los cuales reducimos la incertidumbre que produce nuestra limitación para enfrentarnos a la complejidad de estímulos ambientales”. Por ende, se trata de una forma de pensamiento rápida e intuitiva que otorga estimaciones de probabilidad para la toma de decisiones. Sin embargo, el uso de heurísticos conlleva potenciales errores evitables que pueden conducir a diagnósticos incorrectos⁷ (Ejemplo 1). La

Figura 1. Ilustración de los umbrales diagnóstico y terapéutico.



En rojo se señala el umbral diagnóstico que indica la probabilidad máxima que se tolera para excluir un diagnóstico por ser improbable. En verde se señala el umbral terapéutico que indica la probabilidad mínima a partir de la cual se asume el diagnóstico como probable para iniciar un tratamiento, entre otras decisiones. En naranja queda la probabilidad intermedia en la cual los test diagnósticos son de mayor utilidad.

Dx: diagnóstico.

Fuente: elaboración propia.

Ejemplo 2.

Un niño en edad escolar con fiebre baja, de pocas horas de evolución y sin síntomas orientativos a un foco infeccioso específico, tiene baja probabilidad de tener como foco una infección urinaria (debajo del umbral diagnóstico). Sin embargo, si el mismo paciente presenta además síntomas urinarios, se beneficiaría de exámenes de orina (entre ambos umbrales). Si los exámenes de orina fueran compatibles con una infección urinaria, se iniciaría tratamiento antibiótico (sobre el umbral terapéutico).

medicina basada en la evidencia entrega herramientas para “objetivar” de alguna forma la experiencia clínica, evitar sesgos y facilitar la interpretación de las situaciones clínicas.

Ejemplo 1.

Un profesional de salud, tras examinar a un hombre con un cuadro de dolor abdominal sin signos de alarma y exámenes generales de laboratorio en rangos normales, sospecha un cuadro de síndrome de intestino irritable. Sin embargo, recuerda que hace una semana atendió un paciente con un cuadro clínico similar el cual fue finalmente diagnosticado con porfiria. Por este motivo, decide solicitar exámenes específicos para descartar esa enfermedad. Este pensamiento intuitivo corresponde al heurístico de disponibilidad, el cual se basa en reconocer elementos familiares en situaciones nuevas a partir de la información recordada recientemente (“disponible”).

La información aportada por los métodos diagnósticos incrementan o disminuyen la probabilidad de una condición particular⁸, moviéndose entre el umbral diagnóstico y el umbral terapéutico (Figura 1). El umbral diagnóstico refleja la probabilidad mínima necesaria para considerar plausible una condición particular, mientras que el umbral terapéutico refleja la confianza necesaria en el diagnóstico para iniciar tratamiento. Por debajo del umbral diagnóstico no merece la pena realizar pruebas debido a que la probabilidad diagnóstica es baja⁹. A la inversa, superando el umbral terapéutico el diagnóstico tiene una probabilidad tan elevada que justifica las decisiones terapéuticas². Entre ambos, cuando la probabilidad diagnóstica es intermedia, se requiere realizar más pruebas para conseguir una probabilidad que esté bajo el umbral diagnóstico o sobre el umbral del tratamiento^{2,9}.

la frecuencia relativa de sujetos en los que la prueba acertó su condición de enfermedad, representado con la fórmula:

$$\frac{(\text{verdaderos positivos} + \text{verdaderos negativos})}{(\text{total de sujetos evaluados})}$$

Sin embargo, es importante considerar que una prueba diagnóstica puede ser más exacta para detectar la enfermedad o para detectar a los individuos sanos, volviéndose útiles en distintos escenarios⁹. Una prueba diagnóstica puede ser más exacta para detectar la enfermedad, o serlo para detectar a los individuos sanos.

Asimismo, la exactitud de las pruebas diagnósticas puede representarse con indicadores como sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, cocientes de probabilidades y curvas características operativas del receptor (ROC por sus iniciales en inglés *Receiver Operating Characteristic*). Estos indicadores suelen ser conocidos por la mayoría de los médicos generales. Sin embargo, existe evidencia de que pueden llegar a ser aplicados incorrectamente¹⁰.

La evaluación de la exactitud se realiza comparando la concordancia de los resultados obtenidos de la prueba diagnóstica en evaluación con los de un estándar de referencia en un mismo grupo de pacientes. El estándar de referencia, también llamado estándar de oro, corresponde a un único examen o combinación de métodos (estándar de oro compuesto), que permite establecer de la mejor manera disponible la presencia o ausencia de una determinada condición⁹. A modo de ejemplo, para el diagnóstico de un tromboembolismo pulmonar agudo, el estándar de referencia es la angiografía por tomografía axial computarizada. Si para diagnosticar la misma condición se utilizara la prueba del dímero-D mediante aglutinación por látex, la estimación de la sensibilidad y la especificidad de sus resultados serían a partir de la comparación de estos con el estándar de oro¹¹. El impacto de una prueba diagnóstica hace referencia a cómo y/o cuánto un resultado dado de una prueba diagnóstica incide en el cuidado posterior del paciente¹². Es decir, la evaluación del impacto determina en qué modo la información entregada por el resultado de la prueba, afecta las decisiones terapéuticas y desenlaces clínicos¹³.

Para determinar el impacto de una prueba diagnóstica se debe realizar un estudio prospectivo de seguimiento a corto y largo plazo. Otra alternativa que puede sumarse a la anterior es realizar un estudio retrospectivo que permita monitorear, entre otras cosas, número de test diagnósticos aplicados posteriormente, tiempo de demora hasta lograr el diagnóstico definitivo o tiempo de demora hasta instaurar el tratamiento definitivo. A modo de ilustración de la aplicación clínica del impacto, tenemos que en el caso de un paciente en el que se evidencian lesiones de mal pronóstico en una imagen cerebral en el contexto de un accidente vascular cerebral sin alternativas terapéuticas (quirúrgicas o endovasculares), no afectaría el manejo del paciente el conocer las características detalladas de las lesiones a través de nuevas pruebas diagnósticas¹⁴.

Este artículo corresponde al séptimo de una serie metodológica de trece revisiones narrativas acerca de tópicos generales en bioestadística y epidemiología clínica. Esta revisión explora y resume en un lenguaje amigable, artículos publicados disponibles en las principales bases de datos y textos de consulta especializados. La serie está orientada a la formación de estudiantes de pre y posgrado. Es realizada por la Cátedra de Medicina Basada en la Evidencia de la Escuela de Medicina de la Universidad de Valparaíso, Chile, en colaboración

con el Instituto Universitario Hospital Italiano de Buenos Aires, Argentina, y el Centro Evidencia UC de la Pontificia Universidad Católica de Chile. El objetivo de este manuscrito es abordar los principales conceptos teóricos y prácticos de las pruebas diagnósticas en seres humanos.

Probabilidades y más probabilidades en el razonamiento clínico

Constantemente en la práctica médica se realizan aproximaciones probabilísticas para determinar la probabilidad que tiene un individuo de padecer una condición particular. Este procedimiento es previo a la realización de una prueba diagnóstica. Esta aproximación diagnóstica inicial corresponde a la probabilidad pretest. Esta prueba depende de la valoración subjetiva del clínico ante la presencia o ausencia de hallazgos semiológicos para el diagnóstico de una condición particular de interés^{15,16}. De forma simplificada, significa que ante la ausencia de información relevante adicional se ha aceptado el uso de la prevalencia de la condición en estudio para estimar la probabilidad pretest¹⁵.

Un resultado negativo de una prueba diagnóstica en el contexto de una alta sospecha clínica o probabilidad pretest elevada (Ejemplo 3) al igual que un resultado positivo de una prueba diagnóstica en el contexto de una probabilidad pretest baja (Ejemplo 4), nos harán dudar en primera instancia del resultado de la prueba. Cuando la probabilidad pretest es intermedia, el resultado de la prueba diagnóstica puede modificar el escenario probabilístico incierto para descartar o confirmar la sospecha diagnóstica (Ejemplo 5).

Los problemas que acarrea realizar pruebas cuando no hay incertidumbre

Ejemplo 3. Alta probabilidad pretest con un resultado negativo.

Se presenta a urgencias un niño de siete años con odinofagia, fiebre mayor a 38 grados Celsius, con adenopatías pultáceas y laterocervicales con inflamación y dolor. Su mamá refiere que su hermano mayor de 10 años tuvo una faringitis por *Streptococcus pyogenes* (confirmada) hace menos de cinco días. Al realizarle un test rápido al niño, este resulta negativo. Al ser la probabilidad pretest del paciente tan alta, se puede considerar la posibilidad de que se trate de un falso negativo, es decir, que la prueba no haya detectado una enfermedad que sí existe. En este escenario sería apropiado solicitar el estándar de oro (cultivo faríngeo). Es importante señalar que, ante la imposibilidad de aplicar un test diagnóstico, se podría bajar el umbral terapéutico (Figura 1) e iniciar un tratamiento antibiótico “empírico”.

Ejemplo 4. Baja probabilidad pretest con un resultado positivo.

Un joven de 18 años, sano, con un examen físico normal y sin antecedentes personales o familiares de enfermedad cardiovascular, acude a realizarse una prueba ergométrica graduada como parte de los exámenes de rutina previo a su ingreso al servicio militar obligatorio. Durante la realización de la prueba, el paciente presenta un descenso horizontal del segmento ST de 2 milímetros, en DIII. Como la probabilidad pretest del paciente es muy baja, es razonable pensar que el resultado sea un falso positivo para infarto agudo al miocardio. Esta situación nos plantea que es inadecuado solicitar un test diagnóstico con baja especificidad en un primer momento cuando la probabilidad pretest es muy baja, dado que ante un resultado positivo el paciente deberá ser sometido a otras pruebas más específicas solo para confirmar que se trataba de un falso positivo.

Tests en el área de incertidumbre

Ejemplo 5. Probabilidad pretest intermedia con resultado positivo.

Consulta una paciente mujer de 31 años, sin antecedentes mórbidos, por cuadro de tres meses de evolución de distensión abdominal y dolor tipo cólico asociado a pujo, tenesmo y deposiciones diarreicas mucosas, intermitentemente. En el examen físico solo impresiona distensión abdominal hipogástrica. Tiene antecedentes familiares de enfermedad inflamatoria intestinal y es fumadora. Ante la hipótesis diagnóstica de enfermedad inflamatoria intestinal se solicita calprotectina fecal, cuyo resultado es elevado en 150 microgramos por gramo (la sensibilidad y especificidad para discriminar la enfermedad inflamatoria intestinal del síndrome de intestino irritable varía según el punto de corte de calprotectina y ronda entre 80 y 100% y entre 74 y 100% respectivamente)¹⁷. Dado que la probabilidad luego de aplicar el test se elevó, se solicita ileocolonoscopia más biopsia que muestra patrón de empedrado, úlceras aftosas, fisuras de la mucosa y biopsia compatible con enfermedad de Crohn. Realizado el diagnóstico se instaura tratamiento.

Figura 2. Cambio en conducta clínica luego de implementar una prueba diagnóstica.



El paciente del Ejemplo 5 comienza en la zona de incertidumbre diagnóstica por su sintomatología y sus antecedentes familiares. Al aplicar una prueba diagnóstica de calprotectina fecal, su resultado elevado aumenta la probabilidad del diagnóstico de “enfermedad inflamatoria intestinal” y sobrepasa el umbral terapéutico, donde podría comenzarse un tratamiento.

Fuente: elaboración propia.

¿Cómo medimos la exactitud diagnóstica?

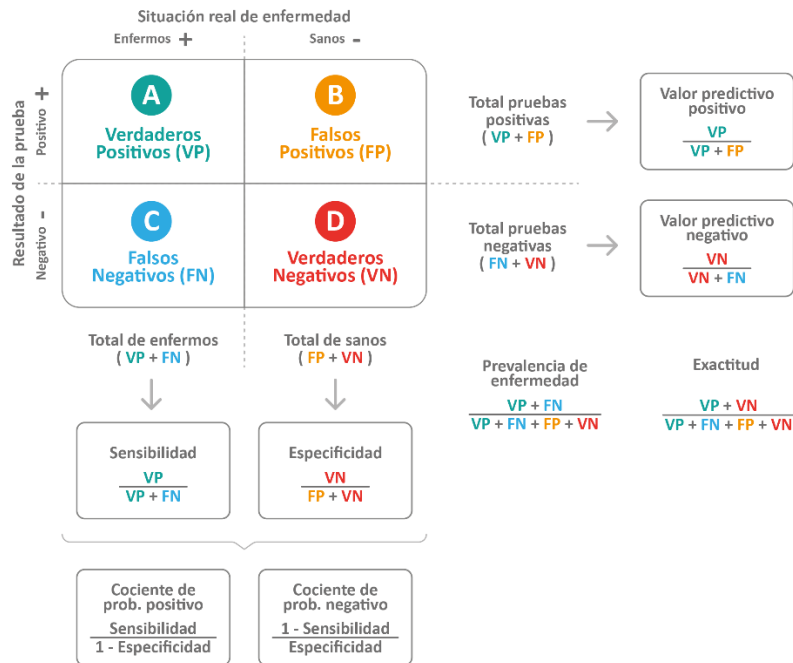
Sensibilidad y especificidad

Al aplicar una prueba diagnóstica existe la posibilidad de clasificar incorrectamente a los individuos sometidos a la misma. Ejemplo de ello son las personas enfermas que en realidad están sanas (falsos positivos) y viceversa, personas sanas cuando en realidad están enfermas (falsos negativos). La información de los valores obtenidos para la prueba, en contraste con los valores del test de referencia o estándar de oro, se presenta en formato de tablas (Figura 3). Las llamadas “tablas de contingencia 2x2” se construyen con dos columnas. Las columnas corresponden al resultado positivo (izquierda) y negativo (derecha) de la condición según el estándar de referencia. A

ellas se suman dos filas que reflejan el resultado positivo (arriba) o negativo (abajo) de la condición, según la prueba índice. Además, se designa una letra a cada celda, siendo de esta manera⁹:

- Verdaderos positivos: aquellos individuos enfermos con resultado positivo del test.
- Falsos positivos: aquellos individuos sanos con resultado positivo del test.
- Falsos negativos: aquellos individuos enfermos con resultado negativo (test).
- Verdaderos negativos: aquellos individuos sanos con resultado negativo (test).

Figura 3. Organización de una tabla de contingencia para la estimación de la precisión diagnóstica.



Fuente: elaboración propia.

Para evaluar las pruebas diagnósticas se utilizan la “sensibilidad” y la “especificidad”². Estos son valores establecidos que se obtienen a partir de la aplicación de la prueba diagnóstica en una población específica al momento de su validación. En este sentido, la sensibilidad y la especificidad son propiedades intrínsecas del test diagnóstico. Sin embargo, su rendimiento también depende de las características de la población en la cual se aplicará. Más adelante en el texto se profundiza sobre estos aspectos¹⁸.

La sensibilidad es la probabilidad de que la prueba clasifique correctamente a los individuos enfermos, o la probabilidad de que el individuo enfermo sea positivo². Las pruebas con alta sensibilidad son útiles en caso de tamizaje, porque tienen muy pocos falsos negativos¹⁹. No obstante, en estos casos también es importante la especificidad para evitar un exceso de falsos positivos, especialmente si estos implican pruebas confirmatorias costosas o invasivas. Además, por la misma baja cantidad de falsos negativos, son especialmente útiles en los casos en que no diagnosticar una enfermedad o evento específico puede resultar peligroso o fatal para los enfermos^{16,18}.

Ejemplo 7.

Una mujer de 67 años se presenta con confusión, náuseas, vómitos y cefalea. Un profesional la evalúa clínicamente y sospecha que puede sufrir un cuadro de hipertensión endocraneana. Dentro de la evaluación neurológica decide realizar un fondo de ojo, dado que la pérdida de la pulsación venosa retinal espontánea es un signo sin falsos negativos para los cuadros de hipertensión endocraneana. Al notar que la pulsación está presente, considera el resultado como un verdadero negativo para hipertensión endocraneana.

La especificidad es la probabilidad de que la prueba clasifique correctamente a los individuos sanos, o la probabilidad de que los individuos sanos tengan un resultado negativo². Una prueba muy específica tiene una tasa de falsos positivos muy baja. Ello le confiere

una capacidad de confirmar una enfermedad muy alta. Esto significa que si un resultado en una prueba muy específica da positivo, hay gran posibilidad de que se trate de un verdadero positivo¹⁸. En la práctica clínica se prefieren pruebas con alta especificidad al momento de confirmar un diagnóstico por su baja cantidad de falsos positivos. Esto es particularmente importante en el caso de enfermedades graves, debido a que un tratamiento oportuno puede disminuir significativamente las consecuencias físicas, económicas y psicológicas¹⁶.

Ejemplo 8.

Una paciente de 27 años con antecedentes familiares de enfermedad de Wilson, se presenta a consulta. Su médico busca evaluar la presencia del anillo de Kayser-Fleisher (anillos dorados en la membrana descemet de la región límbica de la córnea) al realizar el examen físico. Este signo es patognomónico, o sea tiene una especificidad del 100%. Si estuviese presente este anillo podría interpretarlo como la confirmación de la enfermedad, ya que la alta especificidad sugiere que no son probables los falsos positivos.

La estimación de la sensibilidad y la especificidad de una prueba diagnóstica tendrán una mayor aplicabilidad entre más amplia sean las características demográficas y/o clínicas de la muestra de individuos enfermos, como no enfermos de la población donde va a utilizarse el test. Si la muestra es representativa de una población y las estimaciones se utilizan en otra población con características distintas se están dando, y empleando en la práctica, valores de sensibilidad y especificidad equivocados, o por lo menos no aplicables a la población donde se está utilizando.

Como se deduce de lo anteriormente explicado, al ser necesario conocer el estado de salud/enfermedad de los pacientes para poder calcular la sensibilidad y la especificidad, se requiere la contrastación del diagnóstico a través de un método que plantea un parámetro ideal

o estándar de oro (patrón de referencia). Esta es la técnica diagnóstica que define la presencia de la condición con la máxima certeza conocida^{9,19}. Por otra parte, en la práctica clínica habitual los profesionales de la salud se ven enfrentados a pacientes que les consultan con el resultado de una prueba que ya se han realizado. La probabilidad de estar enfermo o no a partir de los resultados del test se conoce como valor predictivo. Este es el tema que se desarrollará a continuación.

Valores predictivos positivos y negativos

El resultado de una prueba diagnóstica conlleva consigo una determinada probabilidad de que dicho resultado categorice correctamente la presencia o ausencia de la una condición, lo que corresponde a los valores predictivos²⁰. El valor predictivo positivo es la probabilidad de que la prueba diagnóstica identifique correctamente los individuos enfermos cuando entrega un resultado positivo. A su vez, el valor predictivo negativo es la probabilidad de que la prueba diagnóstica identifique correctamente los individuos sanos cuando entrega un resultado negativo²¹. Para calcularlos se utilizan proporciones (Figura 3).

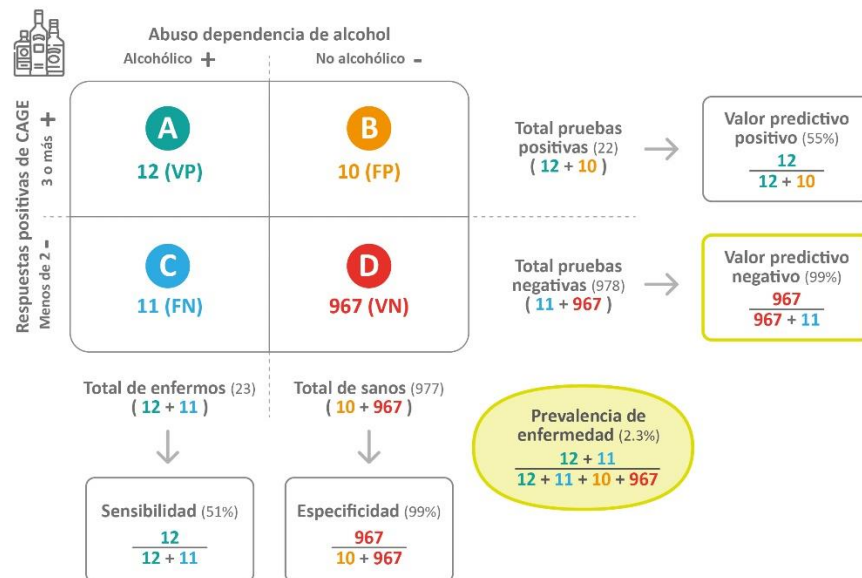
Los valores predictivos están condicionados por la probabilidad a priori de la condición en estudio¹⁸. Cuando la probabilidad a priori es baja, los valores predictivos negativos serán altos y los valores predictivos positivos serán bajos. En este escenario un resultado negativo de una prueba diagnóstica con alto valor predictivo negativo

otorga mayor probabilidad para descartar la condición del paciente que un resultado positivo para confirmarla. Por otro lado, cuando la probabilidad a priori es alta, los valores predictivos positivos serán altos y los valores predictivos negativos serán bajos. En este caso un resultado positivo de una prueba diagnóstica con alto valor predictivo positivo otorga mayor probabilidad para confirmar la condición que un resultado negativo para descartarla^{2,16} (Ejemplos 9A y 9B).

Ejemplo 9-A.

Supongamos que se desea evaluar el abuso o dependencia de alcohol en una población con el cuestionario CAGE (*Cut-down, Annoyed, Guilty, Eye-opener*, cuya sensibilidad del 51% y especificidad del 99% ha sido estimada con anterioridad en los estudios de validación). La localidad A es una comunidad cerrada cuyos valores comunitarios incluyen la abstinencia del alcohol. Si bien no se puede afirmar que nadie toma alcohol, la prevalencia estimada de abuso es baja (23/1000) o (2,3%). Si fijamos nuestra atención en el valor predictivo negativo, este es alto (99%) por la baja prevalencia de la enfermedad. El efecto de la prevalencia sobre el valor predictivo negativo en este escenario está señalado por la baja cantidad de falsos negativos en relación con el total de pruebas negativas. En cambio, el valor predictivo positivo es bajo (55%). Esto indica que es difícil confirmar un diagnóstico con un sólo test en un contexto de baja prevalencia. El efecto de la prevalencia sobre el valor predictivo positivo en este escenario está señalado por la alta cantidad de falsos positivos en relación con el total de pruebas positivas (Figura 4).

Figura 4. Tabla de contingencia de la localidad A.



CAGE: Cut-down, Annoyed, Guilty, Eye-opener.

VP: verdaderos positivos.

FP: falsos positivos.

FN: falsos negativos.

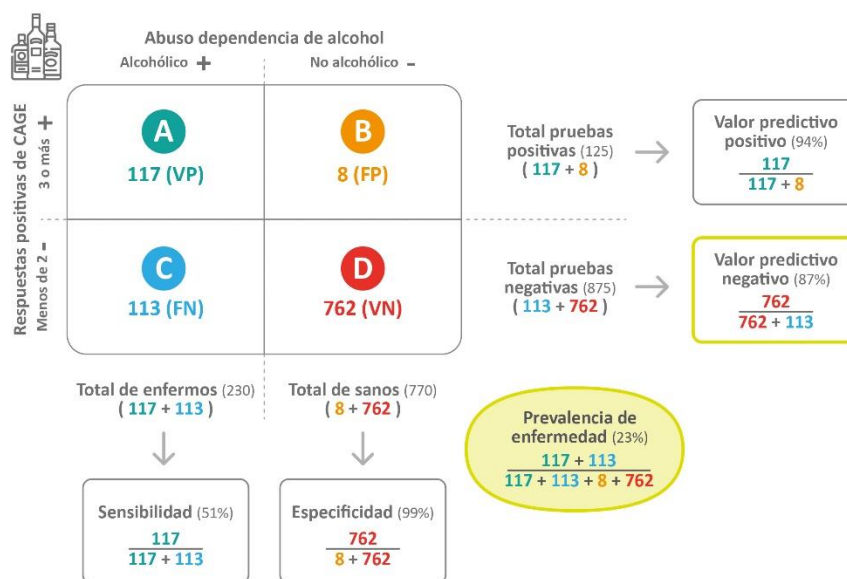
VN: verdaderos negativos.

Fuente: elaboración propia.

Ejemplo 9-B.

El mismo ejemplo pero en otra localidad B, la prueba CAGE tiene los mismos valores de sensibilidad y especificidad ya que estos son propios de la prueba utilizada. Sin embargo, la localidad B tiene mayor consumo de alcohol dado que es una de las principales actividades económicas (producen cerveza), con una prevalencia estimada de abuso o dependencia de un 23%. En este contexto podemos ver que el valor predictivo negativo es más bajo (87%), dado que sería más difícil descartar un diagnóstico en un contexto de alta prevalencia. El efecto de la prevalencia sobre el valor predictivo negativo en este escenario está señalado por la alta cantidad de falsos negativos en relación con el total de pruebas negativas. En cambio, el valor predictivo positivo es más alto (94%) por la alta prevalencia de enfermedad. El efecto de la prevalencia sobre el valor predictivo positivo en este escenario está señalado por la baja cantidad de falsos positivos en relación con el total de pruebas positivas. Por esta razón, un resultado positivo en un contexto de alta prevalencia hace más probable el diagnóstico en comparación a un resultado positivo en contexto de una baja prevalencia de enfermedad²² (Figura 5).

Figura 5. Tabla de contingencia de la localidad B.



CAGE: Cut-down, Annoyed, Guilty, Eye-opener.
 VP: verdaderos positivos.
 FP: falsos positivos.
 FN: falsos negativos.
 VN: verdaderos negativos.
 Fuente: elaboración propia.

Los valores predictivos determinan la probabilidad postest según el resultado de la prueba diagnóstica. Sin embargo, los valores predictivos solo son comparables en poblaciones con una prevalencia o probabilidad pretest de la condición en estudio similar¹⁹.

Cocientes de probabilidades

Los cocientes de probabilidades, también llamados razones de verosimilitud o *likelihood ratio*, comparan la probabilidad de encontrar un determinado resultado (positivo o negativo) de una prueba diagnóstica en personas enfermas respecto a la probabilidad de encontrar ese mismo resultado en personas no enfermas¹⁶. Los cocientes de probabilidades se calculan utilizando la sensibilidad y la especificidad

de una prueba diagnóstica (Figura 3). Los cocientes de probabilidades permiten calcular la probabilidad de una enfermedad posterior a la aplicación de una prueba, adaptándose a las distintas probabilidades previas de estar enfermo en diferentes poblaciones²³.

El cociente de probabilidad positivo determina cuánto más probable es que el resultado de la prueba sea positivo en un paciente enfermo respecto a uno sano. En cambio, el cociente de probabilidad negativo determina cuánto más probable es que el resultado de la prueba sea negativo en un paciente enfermo respecto a uno sano. Para facilitar la interpretación del cociente de probabilidad negativo se utiliza el recíproco del valor calculado para este indicador, cuyo resultado determina cuánto más probable es que el resultado de la prueba sea negativo en un paciente sano respecto a uno enfermo (Ejemplo 10).

Ejemplo 10.

Utilizando los datos de la población de las localidades A y B del Ejemplo 9, podemos calcular el cociente de probabilidad positivo y negativo para el cuestionario CAGE.

$$\text{Cociente de probabilidad positivo} = 0,51/(1-0,99) = 51$$

$$\text{Cociente de probabilidad negativo} = (1-0,51)/0,99 = 0,49$$

El cociente de probabilidad positivo es de 51, lo que significa que un paciente enfermo tiene 51 veces más probabilidades de tener un cuestionario CAGE positivo para alcoholismo en comparación a un paciente sano. El cociente de probabilidad negativo para las localidades A y B es de 0,49 (para calcular su recíproco: $1/0,49 \approx 2$), lo que significa que un paciente sano tiene 2 veces o el doble de probabilidad de tener un cuestionario CAGE negativo para alcoholismo en comparación a un paciente enfermo.

Los cocientes de probabilidad positivos pueden tener valores entre 1 e infinito y los negativos entre 0 y 1. Un cociente de probabilidad de 1 indica nula utilidad para discriminar la presencia o ausencia de una condición²³⁻²⁵ (Tabla 1).

Tabla 1. Potencia diagnóstica.

| Capacidad | Cociente de probabilidad positivo | Cociente de probabilidad negativo |
|-----------|-----------------------------------|-----------------------------------|
| Alta | > 10 | < 0,1 |
| Moderada | 5 a 10 | 0,1 a 0,2 |
| Baja | 2 a 5 | 0,2 a 0,5 |
| Escasa | 1 a 2 | 0,5 a 1 |

Capacidad diagnóstica de una prueba según el valor de cociente de probabilidad.

La primera columna presenta la capacidad o significancia del cociente de probabilidad positivo o negativo para modificar la probabilidad pretest a postest según la magnitud de su valor. Los cocientes de probabilidad mayores a 10 o menores a 0,1 generan grandes cambios de la probabilidad pretest a postest. Estos cocientes serán suficientes en la mayoría de los casos para confirmar (sobre el umbral terapéutico) o descartar la condición en estudio (bajo el umbral diagnóstico)²⁶. Los cocientes de probabilidades de 5 a 10 y de 0,1 a 0,2 generan cambios moderados de la probabilidad pretest a postest. Los cocientes de probabilidad de 2 a 5 y de 0,5 a 0,2 generan cambios pequeños de probabilidad.

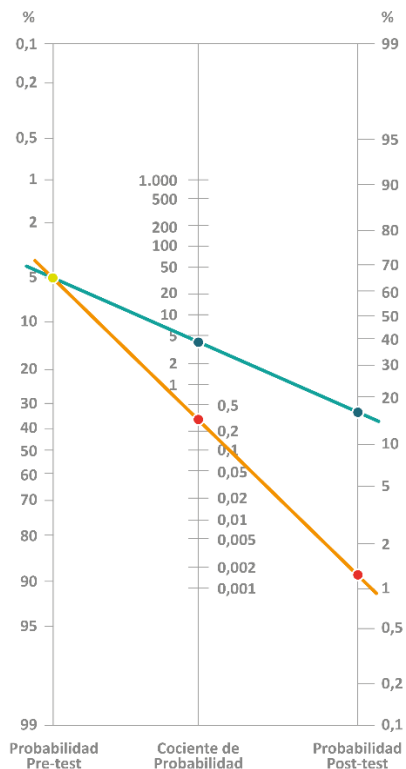
Fuente: elaboración propia.

La forma más práctica y sencilla de interpretar los cocientes de probabilidades es mediante la aplicación del teorema de Bayes con el nomograma de Fagan^{27,28}. En este gráfico la columna izquierda representa la probabilidad pretest, la del centro el cociente de probabilidad de la prueba diagnóstica aplicada y la de la derecha la probabilidad postest¹⁹. Mediante la prolongación de una línea recta que una los valores obtenidos de la primera columna con el de la segunda columna, es posible obtener el resultado de la tercera columna, correspondiente a la probabilidad de tener la condición en función del resultado de la prueba diagnóstica (Ejemplo 11).

Ejemplo 11.

Una paciente de sexo femenino de 85 años consulta por dolor articular matutino en ambas manos de más de una hora de duración. Ante la sospecha de artritis reumatoide, el médico solicita un examen serológico conocido como factor reumatoideo. Para saber la probabilidad de enfermedad de esta paciente se usa el nomograma de Fagan, el cual se visualiza en la Figura 6, una vez que tenga el resultado de la prueba diagnóstica.

Figura 6. Nomograma de Fagan del factor reumatoideo.



La probabilidad pretest definida por el médico del ejemplo corresponde a la prevalencia mundial en mayores de 65 años para esta enfermedad, la cual es de aproximadamente 5% (punto amarillo en la primera columna)²⁹. El cociente de probabilidad positivo es de 4,86 (punto azul en la segunda columna) y el cociente de probabilidad negativo es de 0,38 (punto rojo en la segunda columna)³⁰. La probabilidad posttest para el resultado positivo es de aproximadamente 20% y se obtiene al trazar una línea recta (línea celeste) entre la probabilidad pretest y el cociente de probabilidad positivo. La probabilidad posttest para el resultado negativo es de aproximadamente 1% siguiendo el mismo método (línea naranjada).
Fuente: elaboración propia.

Un resultado positivo para el factor reumatoideo sin otros signos o síntomas que apoyen la presencia de artritis reumatoide, no es suficiente para realizar el diagnóstico y mucho menos justificar un tratamiento³¹.

Curva característica operativa del receptor

Algunas pruebas diagnósticas informan sus resultados en datos continuos u ordinales, tales como la presión arterial o la glicemia. Al utilizar este tipo de datos se debe determinar el punto de corte en donde existe la mayor sensibilidad y especificidad. Es decir, el lugar de la curva en donde mejor se discrimina a los enfermos de los no enfermos³². Sin embargo, no hay un valor que separe claramente a los enfermos de los que no lo están, existiendo valores solapados entre ambos grupos.

Las curvas características operativas del receptor son una representación gráfica que relaciona la proporción de verdaderos positivos (sensibilidad) con la proporción de falsos positivos (1 menos especificidad) para los diferentes posibles valores de una prueba diagnóstica con el propósito de determinar qué valor discrimina mejor entre enfermos y no enfermos. La curva característica operativa del receptor se construye a partir de una gráfica de dispersión, cuyos ejes de ordenadas (y) y abscisas (x) corresponden respectivamente a la sensibilidad y al complemento de la especificidad para los diferentes posibles resultados de la prueba diagnóstica. Desde la esquina inferior izquierda y la esquina superior derecha del gráfico se traza una línea punteada que recibe el nombre de “diagonal de referencia” o “línea de no discriminación”. Esta diagonal de referencia corresponde a la representación teórica de una prueba diagnóstica que no discrimina enfermos de no enfermos (distribución idéntica de los resultados para ambos grupos).

El punto de corte que discrimina mejor entre enfermos y no enfermos dentro de la curva característica operativa del receptor es aquel que alcanza la sensibilidad y especificidad más altas al mismo tiempo. Gráficamente corresponde al punto más cercano al ángulo superior izquierdo de la gráfica, calculado mediante el índice de Youden (sensibilidad + especificidad - 1)³³. Sin embargo, según el objetivo clínico que tenga la prueba diagnóstica, el punto de corte puede ser diferente para así privilegiar la sensibilidad o la especificidad (Ejemplo 12).

Ejemplo 12.

En el siguiente ejemplo, tomado y modificado del libro de Epidemiología Clínica hecho por Feinstein³⁴ se les realizó la prueba ergométrica a una muestra de dos grupos de pacientes, uno con enfermedad coronaria demostrada y otro que no la tiene. Al final de la prueba se les midió el desnivel del segmento ST (Tabla 2).

Tabla 2. Valores obtenidos en la prueba ergométrica.

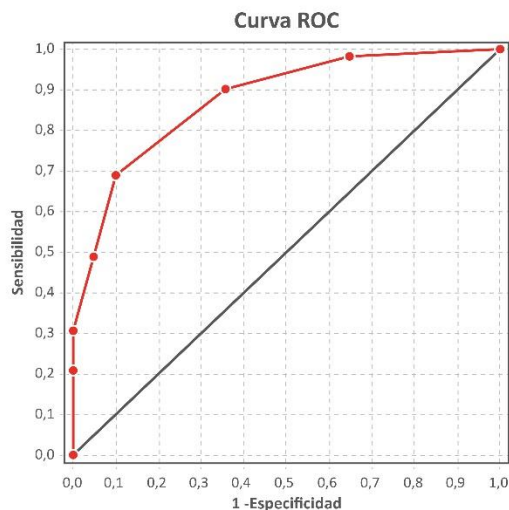
| Desnivel de ST | Con enfermedad coronaria | Proporción acumulada | Sin enfermedad coronaria | Proporción acumulada |
|----------------|--------------------------|----------------------|--------------------------|----------------------|
| ≥ 3 milímetros | 31 | 0,21 | 0 | 0,00 |
| 2,5 - < 3,0 | 15 | 0,31 | 0 | 0,00 |
| 2,0 - < 2,5 | 27 | 0,49 | 7 | 0,05 |
| 1,5 - < 2,0 | 30 | 0,69 | 8 | 0,10 |
| 1,0 - < 1,5 | 32 | 0,90 | 39 | 0,36 |
| 0,5 - < 1,0 | 12 | 0,98 | 43 | 0,65 |
| < 0,5 | 3 | 1,00 | 53 | 1,00 |
| Total | 150 | | 150 | |

De izquierda a derecha, la primera columna señala los rangos de valores de desnivel del segmento ST después de realizada la prueba ergométrica. La segunda columna indica el número de pacientes que fueron clasificados correctamente como enfermos en el rango de valores de desnivel del segmento ST correspondiente. La tercera columna da cuenta de la frecuencia relativa acumulada de pacientes clasificados correctamente como enfermos, es decir la sensibilidad para estos rangos de desnivel del segmento ST. La cuarta columna indica al número de pacientes que fueron clasificados incorrectamente como enfermos en el rango de valores de desnivel del segmento ST correspondiente. Por último, la quinta columna muestra la frecuencia relativa acumulada de pacientes clasificados incorrectamente como enfermos, es decir el complemento de la especificidad para estos rangos de desnivel del segmento ST.

Fuente: adaptación y modificación de Feinstein y colaboradores³⁴.

El punto de corte que mejor discrimina entre enfermos y no enfermos de enfermedad coronaria para esta prueba diagnóstica sería el supradesnivel del segmento ST mayor o igual a 1,5 milímetros, el cual tiene una sensibilidad de 0,69 y una especificidad de 0,90. Sin embargo, en la práctica clínica el punto de corte utilizado para la enfermedad coronaria es el supradesnivel del segmento ST mayor a 1 milímetro que tiene una sensibilidad de 0,90 y una especificidad de 0,64. Este punto de corte privilegia la sensibilidad a expensas de la especificidad^{35,36}, dado que no diagnosticar la enfermedad coronaria cuando existe (falso negativo) puede ser dañino e incluso mortal para los pacientes. Los datos obtenidos en este ejemplo son ilustrados en forma de curva característica operativa del receptor (Figura 7).

Figura 7. Curva característica operativa del receptor de la prueba ergométrica.



La curva característica operativa del receptor de la prueba diagnóstica ergométrica del ejemplo presentado consta de dos ejes que varían de 0 a 1 (0 a 100%), la sensibilidad en el eje vertical (y) y el complemento de la especificidad en el eje horizontal (x). Cada punto rojo en la curva característica operativa del receptor son posibles puntos de corte. La línea diagonal en negro recibe el nombre de diagonal de referencia o línea de no discriminación.

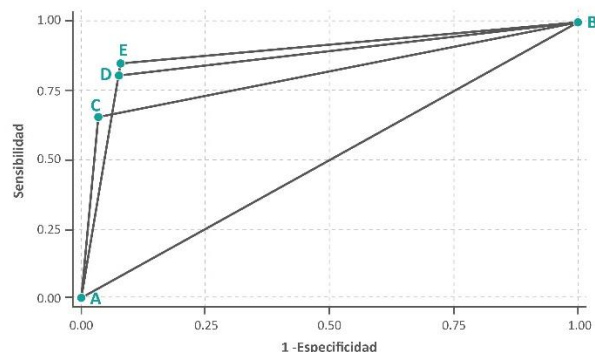
Fuente: adaptación y modificación de Feinstein y colaboradores³⁵.

El área bajo la curva característica operativa del receptor es el indicador global de la precisión de una prueba diagnóstica, cuyo cálculo de obtención escapa del objetivo de este estudio. Esta área oscila entre 0,5 y 1. En el 1 las pruebas diagnósticas alcanzan el 100% de sensibilidad y de especificidad. Un área cercana a 0,5 significa que la prueba diagnóstica no discrimina pacientes enfermos de sanos. El área bajo la curva característica operativa del receptor permite comparar entre dos o más pruebas diagnósticas³⁷, eligiendo en términos generales, la que tenga el mayor área al ser la que mejor discrimine entre enfermos y no enfermos (Ejemplo 12).

Ejemplo 12.

Peonim y colaboradores³⁸ determinaron que la realización en conjunto del antígeno prostático específico y la fosfatasa ácida son el método diagnóstico más exacto para la detección de semen en muestras vaginales humanas. Esta conclusión se basó en la comparación de las curvas características operativas del receptor de ambas pruebas, realizadas tanto de forma separada como en conjunto (Figura 8).

Figura 8. Comparación de curvas características operativas del receptor.



En este gráfico se comparan las curvas características operativas del receptor de la fosfatasa ácida sola (curva ABC), del antígeno prostático específico sólo (curva ABD) y ambas pruebas en conjunto (curva ABE). Las áreas bajo las curvas características operativas del receptor son respectivamente 0,8091, 0,8639 y 0,8823. Por lo tanto, la combinación de ambas pruebas diagnósticas tiene la mayor exactitud diagnóstica.

Fuente: adaptación y modificación de la Figura 1 de Peonim y colaboradores (2013)³⁸.

Conclusiones

Las pruebas diagnósticas asisten a la toma de decisiones clínicas y para su análisis es indispensable entender sus propiedades (sensibilidad, especificidad, valores predictivos y cocientes de probabilidades).

Según el teorema de Bayes, a partir de la probabilidad basal del individuo (probabilidad pretest), las propiedades de la prueba y sus resultados, podemos obtener una nueva probabilidad en relación a la condición en estudio.

Las curvas características operativas del receptor son instrumentos útiles para evaluar pruebas diagnósticas con resultados cuantitativos no dicotómicos, permitiendo discriminar entre dos estados de salud.

La interpretación correcta de los resultados de una prueba puede evitar errores en la toma de decisiones con consecuencias negativas para quienes son sometidos a estas pruebas.

Notas

Roles de autoría

Todos los autores contribuyeron en la planificación y escritura del manuscrito original, así como en la introducción, conceptualización, ejemplos y conclusiones del artículo.

Declaración de conflictos de intereses

Los autores completaron la declaración de conflicto de interés de ICMJE y declararon que no recibieron fondos por la realización de este artículo; no tienen relaciones financieras con organizaciones que puedan tener interés en el artículo publicado en los últimos años y no tienen otras relaciones o actividades que puedan influenciar en la publicación del artículo.

Financiamiento

Los autores declaran que no hubo fuentes externas de financiamiento.

Consideraciones éticas

Este estudio no requirió de la evaluación por parte de un comité de ética, debido a que trabajó sobre fuentes secundarias.

Idioma de envío

Español.

Referencias

1. McGee S. Evidence-Based Physical Diagnosis. 4° Ed. Elsevier; 2017.
2. Molina Arias M. Characteristics of diagnostic tests. Rev Pediatr Aten Primaria. 2013;15(58):169–73. [On line]. | [Link](#) |
3. Cobo E, Muñoz P, González JA. Bioestadística para no estadísticos. 1° Ed. Elsevier Masson; 2007.
4. Gross RD. Making Medical Decisions: An Approach to Clinical Decision Making for Practicing Physicians. 1° Ed. ACP Press; 1999.
5. McDonald CJ. Medical heuristics: the silent adjudicators of clinical practice. Ann Intern Med. 1996 Jan 1;124(1 Pt 1):56–62. | [CrossRef](#) | [PubMed](#) |
6. Pérez Echeverría MP. Psicología del razonamiento probabilístico. 1° Ed. UAM; 1990.
7. Elstein AS. Thinking about diagnostic thinking: a 30-year perspective. Adv Health Sci Educ Theory Pract. 2009 Sep;14 Suppl 1:7–18. | [CrossRef](#) | [PubMed](#) |
8. Araujo Alonso M. Critical analysis of studies of diagnostic tests: I. Medwave. 2012 Aug 1;12(07):e5465–e5465.
9. Bravo-Grau S, Cruz Q JP. Estudios de exactitud diagnóstica: Herramientas para su Interpretación. Rev chil radiol. 2015;21(4):158–64. [On line]. | [Link](#) |
10. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. BMJ. 2002 Apr 6;324(7341):824–6. | [CrossRef](#) | [PubMed](#) |
11. Froehling DA, Elkin PL, Swensen SJ, Heit JA, Pankratz VS, Ryu JH. Sensitivity and specificity of the semiquantitative latex agglutination D-dimer assay for the diagnosis of acute pulmonary embolism as defined by computed tomographic angiography. Mayo Clin Proc. 2004 Feb;79(2):164–8. | [CrossRef](#) | [PubMed](#) |
12. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011 Aug 4;365(5):395–409. | [CrossRef](#) | [PubMed](#) |
13. Carlos R, Gareen I, Gatsonis C, Gorelick J, Kessler L, Lau J, et al. Standards in the Design, Conduct and Evaluation of Diagnostic Testing for Use in Patient Centered Outcomes Research. PCORI. 2012.[On line]. | [Link](#) |
14. Araujo M. Estudios sobre el diagnóstico de las enfermedades. Medwave. 2011 Jul 1;11(07). [On line]. | [Link](#) |
15. Burgos D ME, Manterola D C. Assessment of diagnostic test studies. Rev Chil Cir. 2010;62(3):301–8. [On line]. | [Link](#) |
16. Mark DB, Wong JB. Decision-making in clinical medicine. In: Harrison's Principles of Internal Medicine. 18° Ed. McGraw Hill Professional; 2011:19–29.
17. Vásquez-Morón JM, Argüelles-Arias F, Pallarés-Manrique H, Ramos-Lora M. Utility of fecal calprotectin in inflammatory bowel disease. RAPD. 2017;40(2). [On line]. | [Link](#) |
18. Escrig-Sos J, Martínez-Ramos D, Miralles-Tena JM. Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso. Cirugía Española. 2006 May;79(5):267–73. | [CrossRef](#) |
19. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Investigación clínica II. Estudios de proceso (prueba diagnóstica) [Clinical research II. Studying the process (the diagnosis test)]. Rev Med Inst Mex Seguro Soc. 2011 Mar-Apr;49(2):163–70. | [PubMed](#) |
20. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994 Jul 9;309(6947):102. | [CrossRef](#) | [PubMed](#) |
21. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pitfalls, and Pitfalls in Research and Practice. Front Public Health. 2017 Nov 20;5:307. | [CrossRef](#) | [PubMed](#) |
22. Simel D, Rennie D. The Rational Clinical Examination: Evidence-Based Clinical Diagnosis. McGraw-Hill; 2008.
23. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ. 2004 Jul 17;329(7458):168–9. | [CrossRef](#) | [PubMed](#) |
24. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980 May 15;302(20):1109–17. | [CrossRef](#) | [PubMed](#) |
25. McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002 Aug;17(8):646–9. | [CrossRef](#) | [PubMed](#) |
26. Capurro D, Rada G. The diagnostic process. Rev Med Chil. 2007 Apr;135(4):534–8. [On line]. | [Link](#) |
27. Fagan TJ. Letter: Nomogram for Bayes theorem. N Engl J Med. 1975 Jul 31;293(5):257. | [CrossRef](#) | [PubMed](#) |
28. Aznar-Oroval E, Mancheño-Alvaro A, García-Lozano T, Sánchez-Yepes M. Razón de verosimilitud y nomograma de Fagan: 2 instrumentos básicos para un uso racional de las pruebas del laboratorio clínico [Likelihood ratio and Fagan's nomogram: 2 basic tools for the rational use of clinical laboratory tests]. Rev Calid Asist. 2013 Nov-Dec;28(6):390–1. | [CrossRef](#) | [PubMed](#) |
29. Artritis reumatoide. Universidad Católica de Chile. [On line]. | [Link](#) |
30. Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, et al. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. Ann Intern Med. 2007 Jun 5;146(11):797–808. | [CrossRef](#) | [PubMed](#) |

31. Nicoll D, Lu CM, McPhee SJ. Guide to Diagnostic Tests. 7^o Ed. McGraw-Hill Education; 2017.
32. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr.* 2007 May;96(5):644-7. | [CrossRef](#) | [PubMed](#) |
33. Böhning D, Böhning W, Holling H. Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. *Stat Methods Med Res.* 2008 Dec;17(6):543-54. | [CrossRef](#) | [PubMed](#) |
34. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research.* 2^o Ed. W.B. Saunders Company; 1985.
35. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem.* 1993 Apr;39(4):561-77. | [PubMed](#) |
36. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ.* 1994 Jul 16;309(6948):188. | [CrossRef](#) | [PubMed](#) |
37. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988 Sep;44(3):837-45. | [PubMed](#) |
38. Peonim V, Worasuwanarak W, Sujirachato K, Teerakamchai S, Srisont S, Udnoon J, et al. Comparison between prostate specific antigen and acid phosphatase for detection of semen in vaginal swabs from raped women. *J Forensic Leg Med.* 2013 Aug;20(6):578-81. | [CrossRef](#) | [PubMed](#) |

Correspondencia a
Angamos 655, Reñaca,
Viña del Mar



Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.