

Conceptos Sobre Pruebas Diagnósticas y su Aplicación

Concepts Regarding Diagnostic Tests and their Application

Carlos Manterola^{1,2}; Sofía Bendel³; Josué Rivadeneira^{2,4} & Juan Pablo Holguín^{2,5}

MANTEROLA, C.; BENDEL, S.; RIVADENEIRA, J. & HOLGUÍN, J.P. Conceptos sobre pruebas diagnósticas y su aplicación. *Int. J. Morphol.*, 43(5):1785-1795, 2025.

RESUMEN: Con el permanente avance tecnológico, las pruebas diagnósticas (PD) continúan mejorando, presentándonos nuevas alternativas. Dada la gran cantidad de PD en uso, es necesario realizar evaluaciones periódicas de las diversas opciones existentes, de forma tal de poder determinar su rol en la práctica clínica. En primer término, se ha de determinar la precisión de la PD, incluyendo su sensibilidad y especificidad, sus valores predictivos, las razones de verosimilitud, y las curvas de característica operativa del receptor (ROC). Luego, se debe determinar su papel en alguna vía diagnóstica, para posteriormente examinar su efecto en el resultado de los enfermos. En este manuscrito delinearemos algunos principios respecto de estándar de referencia, propiedades y poder discriminatorio de una PD (validez y seguridad de una PD, razones de verosimilitud, probabilidades pre y post-test, curvas ROC, sesgos más frecuentes en estudios de PD, niveles de evidencia y grados de recomendación de estudios de PD; aplicación de listas de verificación y de escalas de calidad metodológica para estudios de PD; y conceptos sobre el uso de PD múltiples y aplicación de estrategias de tamizaje. El objetivo de este manuscrito fue generar un documento de estudio para revisar y discutir características esenciales de las PD; proporcionando algunos ejemplos para su mejor comprensión.

PALABRAS CLAVE: Diagnóstico; Pruebas diagnósticas de rutina; Técnicas y procedimientos de diagnósticos; Precisión diagnóstica; Precisión de pruebas diagnósticas; Propiedades de pruebas diagnósticas; Sensibilidad; Especificidad; Valor predictivo; Razón de verosimilitud; Curva característica operativa del receptor.

INTRODUCCIÓN

Las pruebas diagnósticas (PD), constituyen un pilar de la práctica profesional, dentro del proceso diagnóstico-terapéutico cotidiano, pues son útiles para plantear diagnósticos, pronósticos y terapéuticos permitiendo tomar una decisión informada (Burgos & Manterola, 2010).

Para una importante proporción de los clínicos, algunos términos como sensibilidad, especificidad, valores predictivos, curvas receptor-operador etc., constituyen conceptos abstractos. Sin embargo, cada día se hace más necesario comprender estos conceptos, pues nos permite entre otras cosas un mejor entendimiento entre pares y con quienes realizan PD, desde el laboratorio básico y la anatomía patológica, hasta la generación de las imágenes más sofisticadas; así como el entender mejor su real utilidad, permite una mejor comunicación con los pacientes, en especial cuando se trata de explicar porque

se requiere hacer tal o cual prueba y que significa tal o cual resultado (Vali *et al.*, 2021).

Una situación común en la práctica clínica es tener que decidir cuándo una PD es normal, ya que de esto puede depender la instauración de un tratamiento, la indicación de una cirugía, e incluso el pronóstico de un paciente. Pero, las definiciones de normalidad pueden ser distintas. Es así como, si se enfoca desde la distribución normal de Gauss, se considerará normalidad a aquella medida que represente el promedio ± 2 desviaciones estándar (lo que supone asumir que la distribución es normal y que todas las anomalías tienen la misma frecuencia de aparición). Otra forma de entender la normalidad es identificando a quienes se comportan en forma extrema (“outliers”), y a quienes quizás no se puede ofrecer prevención o tratamiento. La aceptación más empleada en relación con

¹ Universidad de La Frontera, Centro de Estudios Morfológicos y Quirúrgicos (CEMyQ), Chile.

² Universidad de La Frontera, Programa de Doctorado en Ciencias Médicas, Chile.

³ Universidad San Sebastián, Santiago, Chile.

⁴ Zero Biomedical Research, Quito, Ecuador.

⁵ Universidad del Azuay, Cuenca, Ecuador.

FUNDED. Parcialmente financiado por proyecto DI19-0030, Universidad de La Frontera, Chile.

PD es la que considera normal al intervalo de resultados fuera del cual, la existencia de la enfermedad objeto de estudio resulta altamente probable (Chassé & Fergusson, 2019).

Sin embargo, una PD puede ser utilizada en el proceso diagnóstico de una enfermedad, condición o evento de interés (EI), así como también en escenarios de tamizaje (screening), y de diagnóstico diferencial. Por otra parte, los elementos necesarios para definir las propiedades de una PD son determinar el EI a estudiar, el estándar de referencia (ER), y la población a la que se aplicará la PD.

En este manuscrito delineamos algunos principios respecto de ER, propiedades y poder discriminatorio de una PD, validez de una PD (sensibilidad y especificidad); seguridad de una PD (valores predictivos); probabilidad pre-test (prevalencia), razones de verosimilitud, puntos de corte de una PD; curvas ROC, probabilidad post-test; sesgos en estudios de PD; niveles de evidencia y grados de recomendación de estudios de PD; la aplicación de listas de verificación para el reporte de resultados (QUADAS-2, GRRAS, QAREL y TRIPOD) y de escalas de calidad metodológica para estudios de PD (Arrivé, STARD y MInCir-Dg); conceptos sobre la utilización de PD múltiples (en serie y en paralelo); y sobre la de aplicación de estrategias de tamizaje.

El objetivo de este manuscrito fue generar un documento de estudio para revisar y discutir características esenciales de las PD; proporcionando algunos ejemplos para su mejor comprensión.

ASPECTOS HISTÓRICOS

El proceso diagnóstico tiene sus orígenes en los médicos griegos, quienes obtenían una historia clínica detallada y practicaban la auscultación directa. Eran maestros de la observación, tanto así que sus descripciones de los pacientes podrían encajar perfectamente en los textos modernos sin demasiadas modificaciones (Walker, 1990) (Tabla I).

En 1666, Thomas Sydenham desarrolló la nosología (rama de la medicina que se ocupa del concepto, definición, clasificación y nomenclatura de las enfermedades), lo que fue fundamental para la evolución del diagnóstico (una enfermedad puede causar manifestaciones diferentes de un individuo a otro).

En 1761, Morgagni fundó la anatomía patológica (a pesar de que la primera autopsia forense conocida se realizó en 1302 en Bolonia). Comenzaron las disecciones con fines educativos de cuerpos de criminales ejecutados (*Anatomia publica*), y de cuerpos de personas que morían de enfermedades (*Anatomia privata*). También durante 1761, Leopold Auenbrugger descubrió la percusión clínica, que posteriormente popularizó Jean Corvisart en 1808 (Berger, 1999a).

Durante 1816, Laennec descubrió la auscultación y el estetoscopio; encontrando además, la confirmación del diagnóstico clínico en la mesa de autopsias (Berger, 1999b).

Tabla I. Algunos avances cruciales en la historia del diagnóstico físico.

Fecha	Autor	Desarrollo
1666	Sydenham	Nosología. Fundamental en la evolución del proceso diagnóstico
1761	Morgagni	Comienza la patología
1761	Auenbrugger	Descubrimiento de la percusión
1816	Laennec	El estetoscopio y la auscultación
1850	Helmholtz	Oftalmoscopio
1854	García	Laringoscopio
1871	Wunderlich	Termómetro
1875	Erb y Westphal	Martillo de reflejos
1895	Röntgen	Rayos X
1897	Killian	Broncoscopio rígido
1903	Curie	Radio y Polonio (equipos de rayos X)
1910	Jacobaeus	Toracoscopia
1948	Mainini	Prueba de embarazo
1949	Ludwig	Ultrasonido
1962	Holmes, Wright y Meyerdirk	Ecógrafo modo B
1968	Ikeda	Broncoscopio flexible
1970	Housfield	Primera imagen de tomografía computada
1973	Lauterbur	Primera imagen de resonancia magnética

En 1850, Helmholtz desarrolló el oftalmoscopio. En 1871, Carl Wunderlich inventó el termómetro. En 1875, Erb y Westphal desarrollan el martillo de reflejos, para el estudio de los reflejos de los tendones profundos; el que vino a complementar el examen neurológico de la época, que consistía en el reflejo pupilar de la luz; la ausencia del reflejo de parpadeo en la parálisis de Bell; el reflejo de agarre; y la flexión de una pierna espástica al acariciar la planta del pie (Berger, 1999a,b). En 1895, Röntgen descubrió los rayos X.

CONCEPTOS ASOCIADOS AL ESTUDIO DE PDs

Prueba diagnóstica: Acción o grupo de acciones dirigidas a confirmar o descartar la presencia o ausencia de una determinada enfermedad, condición o EI. Estas acciones incluyen procesos básicos como la anamnesis y el examen físico, hasta más avanzados como pruebas de biología molecular, y exámenes de imagen sofisticados.

Estándar de referencia. Corresponde al denominado “gold standard” o “prueba de oro”. Es la PD más cercana a la veracidad del fenómeno en estudio, de la que se puede disponer en un momento determinado. Es decir, hasta el momento del estudio, cual es la mejor alternativa diagnóstica existente para estudiar una determinada enfermedad, condición, o EI; por ende, la mejor opción para identificar positivos y negativos de dicha enfermedad, condición, o EI.

Categorías de clasificación. Corresponden a las opciones de clasificación binaria en las que el resultado de una PD indica correcta o incorrectamente la presencia de un EI, condición, o enfermedad; y pueden ser (Fig. 1).

- Verdadero Positivo (VP): Representa la identificación correcta de un EI, condición, o enfermedad por parte de la PD en evaluación; lo que significa, que la PD ha detectado e identificado correctamente el EI.
- Verdadero Negativo (VN): Representa a la determinación correcta de que no existe el EI, condición, o enfermedad por parte de la PD en evaluación; lo que significa, que la PD ha detectado e identificado correctamente la no existencia del EI.
- Falso Positivo (FP): Es la identificación incorrecta del EI por parte de la PD en estudio; lo que significa, que la PD ha detectado algo que en realidad no es el EI.
- Falso Negativo (FN): Corresponde a la falta de identificación del EI por parte de la PD en estudio; lo que significa, que la PD no ha detectado el EI.

En las pruebas de hipótesis estadísticas, estos conceptos se conocen como errores de tipo I y tipo II. Un error de tipo I (FP), ocurre cuando el investigador rechaza una hipótesis nula que en realidad es verdadera en la

población; un error de tipo II (FN), ocurre si el investigador no rechaza una hipótesis nula que en realidad es falsa en la población (Shreffler & Huecker, 2024). Aunque ambos nunca se pueden evitar, se puede reducir su probabilidad de ocurrencia incrementando el tamaño de la muestra (a mayor muestra, menor probabilidad de diferir sustancialmente de la población que le dio origen).

Resultado de PD en estudio	Estado respecto a la enfermedad o el EI según el ER	
	Presente	Ausente
Positivo	a (enfermos c/ PD positiva) = VP	b (no enfermos c/ PD positiva) = FP
Negativo	c (enfermos c/ PD negativa) = FN	d (no enfermos c/ PD negativa) = VN

Fig. 1. Tabla tetracórica para tabular datos en un estudio de PD. VP: verdadero positivo (enfermos con PD positiva); FP: falso positivo (no enfermos con PD positiva); FN: falso negativo (enfermos con PD negativa); VN: verdadero negativo (no enfermos con PD negativa).

PROPIEDADES DE UNA PD

Poder discriminatorio de una PD. Una de las características fundamentales de una PD es su poder discriminatorio, y éste, tiene relación con la variabilidad de la prueba, la reproducibilidad de los hallazgos, y la variabilidad de la población sana, o la determinación de los rangos de valores normales de la prueba en cuestión.

Los criterios de evaluación de una PD son la validez (sensibilidad y especificidad) y la seguridad (los valores predictivos). Estos, proporcionan información acerca de la capacidad de discriminación de la PD, pues son marcadores de la proporción de enfermos (con EI) y no enfermos (sin EI), que son clasificados correctamente.

Validez de una PD. Las características intrínsecas de las PD se evalúan mediante la determinación de sensibilidad y especificidad.

Sensibilidad (S). Es la proporción de sujetos que, teniendo el EI en estudio, definido por el ER, éste es identificado por la PD en evaluación. Dicho de otra forma, la S representa a los sujetos con el EI que presentan un resultado positivo en la PD en evaluación (Fig. 2). Es decir, se relaciona con el concepto “positivo para enfermedad o EI positivo”. La S responde al enunciado: si el paciente tiene realmente la enfermedad, condición en estudio e EI, ¿cuál es la probabilidad de que la PD sea positiva? (Manterola, 2009).

De tal modo que las PD sensibles serán positivas en presencia del EI; razón por la cual son particularmente útiles cuando se requiere realizar tamizajes, cuando el precio a pagar por no detectar el EI es muy alto (por ej.: HIV, cáncer). Del mismo modo, son útiles en las primeras etapas de un plan de estudio, cuando se requiere descartar enfermedades improbables; y también cuando la probabilidad de ocurrencia del EI es baja y el propósito de la prueba es descubrirla.

Especificidad (E). Corresponde a la proporción de sujetos libres de la enfermedad, condición o EI en estudio, definida por el ER, a los que la PD en evaluación identifica como no enfermos o sin EI. Dicho de otra forma, la E representa a los sujetos sin el EI que presentan un resultado negativo en la PD en evaluación (Fig. 2). En consecuencia, las PD específicas son útiles cuando su resultado es positivo; son de gran utilidad para confirmar un diagnóstico que ha sido sugerido por otras PD; rara vez son positivas en ausencia del EI; tienen un rol importante en situaciones en las que los falsos positivos de una PD pueden dañar al paciente. Es decir, se relaciona con el concepto “negativo para enfermedad o EI negativo”. La E responde al enunciado: si el sujeto no tiene la enfermedad, condición o EI, ¿cuál es la probabilidad de que la PD sea negativa? (Manterola, 2009).

Para calcular la S y E de una PD se deben seguir los siguientes pasos: definir el ER; determinar los grupos de sujetos a estudio (uno con y otro sin EI, definido por el ER); categorizar a los sujetos en estudio como positivos o

negativos para el EI en estudio, según la PD en evaluación, clasificar a cada sujeto como positivo o negativo; identificar los sujetos que son VP, VN, FP y FN; y luego, aplicando las definiciones de S y E, calcular los valores correspondientes, para lo que será indispensable construir tablas tetracóricas, de contingencia, o de 2 x 2 (Fig. 1).

Seguridad de una PD. Las características extrínsecas de las PD se determinan mediante:

- Valor predictivo positivo (VPP). Corresponde a la proporción de sujetos que con una PD positiva para EI que están realmente enfermos. El VPP de una PD se puede explicar con la siguiente condición: si el resultado de una PD es positivo ¿qué probabilidad tiene el sujeto de presentar el EI en estudio? La fórmula para calcular esta propiedad de una PD se resume en la Figura 2 (Manterola, 2009).
- Valor predictivo negativo (VPN). Corresponde a la proporción de sujetos que con una PD negativa para un determinado EI, no tienen el EI. Así, esta propiedad se puede explicar con la siguiente situación: si el resultado de una PD es negativo ¿qué probabilidad tiene el sujeto de no presentar el EI? La fórmula para calcular esta propiedad de una PD se resume en la Figura 2 (Manterola, 2009).

Al igual que para el VPP, el VPN depende de la prevalencia del fenómeno en evaluación en la población en estudio.

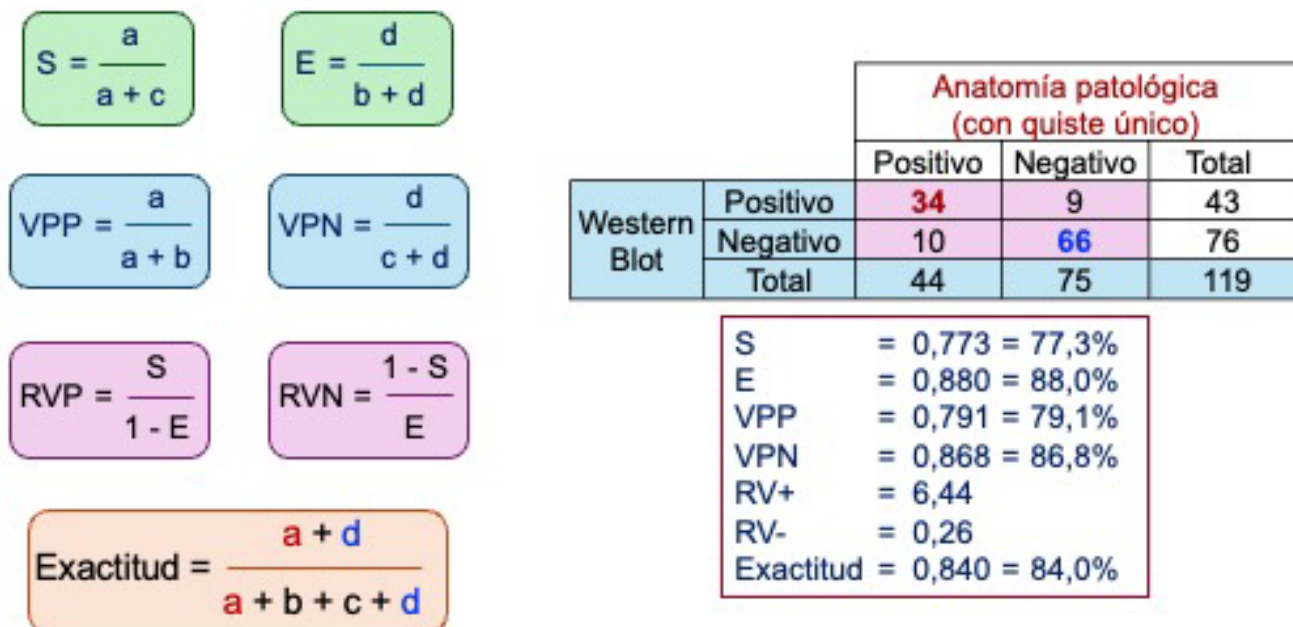


Fig. 2. Fórmulas para calcular sensibilidad, especificidad, valores predictivos, razones de probabilidad y exactitud de una PD. Ejemplo del resultado del cálculo de estas en un estudio de PD realizado para determinar exactitud de algunas PD (en este caso de Western Blot) en pacientes con equinococosis hepática, en una muestra de 119 sujetos (44 con y 75 sin el EI según el ER). Datos obtenidos de Manterola *et al.* (2005).

Prevalencia del resultado. A pesar de que S y E definen completamente la validez de una PD, no proporcionan información relevante a la hora de la toma de decisiones ante un determinado resultado de la PD. No obstante, son las propiedades intrínsecas de la PD y por lo tanto, definen su validez independientemente de cuál sea la prevalencia de la condición, enfermedad o EI en la población a la cual se aplica (probabilidad pretest o probabilidad estimada antes de la aplicación de una PD); es decir, las propiedades intrínsecas no serán modificadas por la prevalencia o probabilidad del EI, por lo cual sus resultados serán extrapolables a otras poblaciones.

Por otra parte, los valores predictivos son de gran utilidad para la toma de decisiones y para explicar a pacientes y familiares información sobre su diagnóstico. Sin embargo, presentan la limitación de depender de la frecuencia del EI en estudio en la población objeto de estudio. Es decir, cuando la prevalencia de un EI es baja, un resultado negativo permitirá descartar el EI con mayor seguridad; determinando así un mayor VPN. Por el contrario, un resultado positivo no permitirá confirmar el diagnóstico, resultando un bajo VPP (Manterola, 2009).

De tal forma, la prevalencia del EI en estudio influye en los valores predictivos de una PD; por ende, no deben ser utilizados como índices para comparar PD diferentes, como tampoco para extrapolar resultados de otros estudios en datos de un centro en particular.

Por lo anteriormente expuesto es indispensable recurrir a otros índices que sean clínicamente útiles y no dependan de la prevalencia del EI en la población a estudio; contexto en el que corresponde aplicar las razones de probabilidad (RP), razones de verosimilitud (RV), cociente de probabilidades; o “Likelihood ratios” (LHR).

Razones de verosimilitud, razones de probabilidad, “Likelihood ratios” (RV, RP, o LHR). Constituyen un método para evaluar el rendimiento de una PD y ayudar en la toma de decisiones clínicas. Aunque la S y E son útiles para pruebas binarias, no pueden aplicarse directamente a pruebas con más de dos posibles resultados; en las que las RV, RP, o LHR son muy útiles. También son adecuadas para PD con resultados continuos (Ost, 2022).

Estos, miden cuánto más probable es un resultado concreto (positivo o negativo) según la presencia o ausencia de un EI. Calcular las RP, RV o LHR, permite obtener mayor precisión en la información de una PD; y tienen tres ventajas interesantes: son intuitivas, simplifican el cálculo de los valores predictivos, y de la evaluación general de pruebas secuenciales. Sus desventajas son la no linealidad, y la

necesidad de recalcular las probabilidades en forma de odds (Timsit *et al.*, 2018).

Los LHR se obtienen combinando S y E de una PD; y pueden ser positiva o negativa. Indican cuánto más probable es un determinado resultado de una PD en un sujeto con un EI dado comparado con un sujeto sin el EI (Elston, 2022).

Las RP, RV o LHR, proporcionan la ventaja de relacionar S y E en un solo índice. Además, se puede obtener RP, RV o LHR según distintos niveles de una nueva medida y no es necesario expresar la información de forma dicotómica, como resultado de normal o anormal o bien positivo y negativo (Fig. 2).

Asimismo, no varían con la prevalencia (al igual que lo que ocurre con S y E), lo que permite emplearlos como elementos de comparación entre distintas PD para un mismo diagnóstico, pues se mantienen constantes independiente de la prevalencia del EI en quienes se aplica la PD.

Por otra parte, las RP, RV o LHR son útiles en la práctica clínica debido a que permiten comprender de mejor forma los resultados de una PD (con qué fuerza el resultado positivo de una PD indica la presencia real del EI y la fuerza de un resultado negativo para descartar el EI). De este modo, las RP, RV o LHR indican como el resultado de una PD hace cambiar la probabilidad pre-test a la probabilidad post-test del EI.

-RP, RV o LHR positivo: describe la probabilidad de tener el EI respecto de no tenerlo, teniendo un resultado positivo de la PD en estudio. Corresponde a la relación entre el porcentaje de enfermos que presentan una PD positiva y el porcentaje de “no enfermos” que presentan una PD positiva. Se calcula dividiendo la probabilidad de un resultado positivo en los sujetos con EI entre la probabilidad de un resultado positivo en los sujetos sin EI, y pueden tener un valor entre 1 e infinito. Es, en definitiva, el cociente entre la fracción de VP (sensibilidad) y la fracción de FP (1-especificidad); o la relación entre S y el complemento de la E (Fig. 2).

-RP, RV o LHR negativo: describe la probabilidad de no tener el EI respecto de tenerlo, con un resultado negativo de la PD en evaluación. Corresponde a la relación entre el porcentaje de sujetos con EI y PD negativa y el porcentaje de sujetos sin el EI y PD negativa. Se calcula dividiendo la probabilidad de un resultado negativo en presencia de EI con la probabilidad de un resultado negativo en ausencia de esta. Es decir, corresponde al cociente entre la fracción de FN (1-S) y la fracción de VN (E); o en otras palabras, constituye la relación entre el complemento de S y E (Fig. 2).

Un ejemplo del cálculo de S, E, VPP, VPN, RVP y RVN se puede apreciar en la Figura 2. Se trata de un estudio de PD realizado para determinar la exactitud de algunas PD (en este caso de Western Blot), en pacientes con equinococosis hepática (Manterola *et al.*, 2005). Con una muestra de 119 sujetos (44 con y 75 sin el EI según el ER), se verificó una exactitud de 84%, E y VPP moderados, S y VPN bajos; RV positiva y negativa de 6,44 y 0,26 respectivamente (Fig. 2). Lo que representa que, si la PD es positiva, la probabilidad de tener equinococosis hepática es 6,44 veces mayor en oposición a no tenerla; así como si la PD es negativa, la probabilidad de no tener equinococosis hepática es sólo de 0,25 veces.

Puntos de corte de una PD. Son los valores de una variable continua que permiten diferenciar lo normal de lo anormal; por lo cual, estos valores dependen del objetivo de la PD. Permiten obtener una clasificación correcta (positivos y negativos) de la mayor proporción de sujetos en estudio. De este modo, un punto de corte altamente sensible permite pesquisar la mayor cantidad de enfermos, o sujetos con EI positivo; y un punto de corte altamente específico, permite detectar la mayor cantidad de no enfermos, o sujetos con EI negativo. Para ello, se debe recurrir a la confección de curvas Característica Operativa del Receptor (Receiver Operator Characteristic), también conocidas como curvas ROC.

Curvas de características operativas del receptor (curvas ROC)

Cuando los resultados de una PD son valores continuos y requieren de conversión e interpretación a una forma dicotómica para determinar la presencia de un EI, el método más utilizado es la confección de curvas característica-operativa del receptor (ROC); las que se utilizan para valorar el rendimiento general de una PD, determinar su validez y comparar el rendimiento de dos o más PD. Además, se puede utilizar para seleccionar un valor de corte óptimo para determinar la presencia o ausencia de EI (Nahm, 2022).

Las curvas ROC constituyen una forma de expresar la relación inversa entre S y E. Se traza mediante gráfica de la tasa de verdaderos positivos (S) y la tasa de falsos positivos (1-E), en la que los valores de los ejes van entre 0 y 1. Es ideal para estudiar una PD con resultados en escala continua, y que requieren determinar puntos de corte para clasificar de forma dicotómica positividad y negatividad (Burgos & Manterola, 2010).

Otras de las características de las curvas ROC son: la pendiente de la curva ROC en un intervalo de resultados de la PD, corresponde a la RV o LHR para ese intervalo. El

punto de corte óptimo para considerar como positiva a una PD, no solo depende de la forma de la curva, sino también de la probabilidad pre-test y de los daños relativos de los falsos positivos y falsos negativos. El área bajo la curva ROC (ABC), sólo mide discriminación, no la precisión de las probabilidades. El ABC no representa una buena medida de discriminación si la pendiente de la curva ROC no disminuye de forma equilibrada. El ABC puede aumentar al incluir un gran número de sujetos identificadas correctamente como de muy bajo riesgo para el EI.

La mayor exactitud de una PD se traduce en un desplazamiento “hacia arriba y a la izquierda” de la curva (ABC); lo que sugiere que el ABC se puede emplear como un índice de exactitud global de una PD; donde la exactitud máxima de una PD correspondería a un valor ABC de 1 (100%); y la mínima a uno de 0,5 (50%); lo que se puede apreciar a modo de ejemplo en la Figura 3, con el comportamiento de 3 curvas ROC. La primera, de una escala síntomas de reflujo gastroesofágico, para diagnosticar la presencia de enfermedad por reflujo gastroesofágico, con un ABC de 0,9806 (Manterola *et al.*, 2002); la segunda, de la misma escala síntomas, para diagnosticar la presencia de esofagitis por reflujo, con un ABC de 0,7819 (Torres-Quevedo *et al.*, 2009); y la tercera de la misma escala de síntomas, para diagnosticar la presencia de cáncer de esófago, con un ABC de 0,499 (Fig. 3).

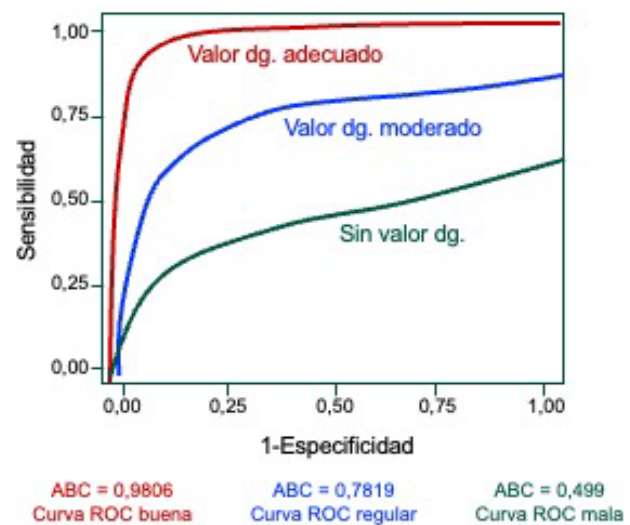


Fig. 3. Tipos de curvas ROC. Distribución de un punto de corte de diferentes curvas ROC. La primera corresponde al punto de corte 3 de una escala de válida y confiable de síntomas de reflujo gastroesofágico, para diagnosticar enfermedad por reflujo gastroesofágico, que determina un ABC de 0,9806 (Manterola *et al.*, 2002); la segunda, al punto de corte 7 de la misma escala síntomas, para diagnosticar esofagitis por reflujo, que determina un ABC de 0,7819 (Torres-Quevedo *et al.*, 2009); y la tercera, que representa una curva ROC sin valor diagnóstico, porque su ABC es menor de 0,50.

Por otra parte, el ABC o c-estadística, se puede utilizar para comparar la bondad de ajuste de los modelos de regresión logística de las curvas ROC. El ABC de una curva ROC se utiliza para cuantificar la capacidad general de una PD para discriminar entre dos resultados, o la capacidad general de una PD para identificar correctamente lo normal versus lo anormal. En combinación, una curva ROC y su ABC correspondiente son funciones de la S y E de un modelo de predicción (Carter *et al.*, 2016).

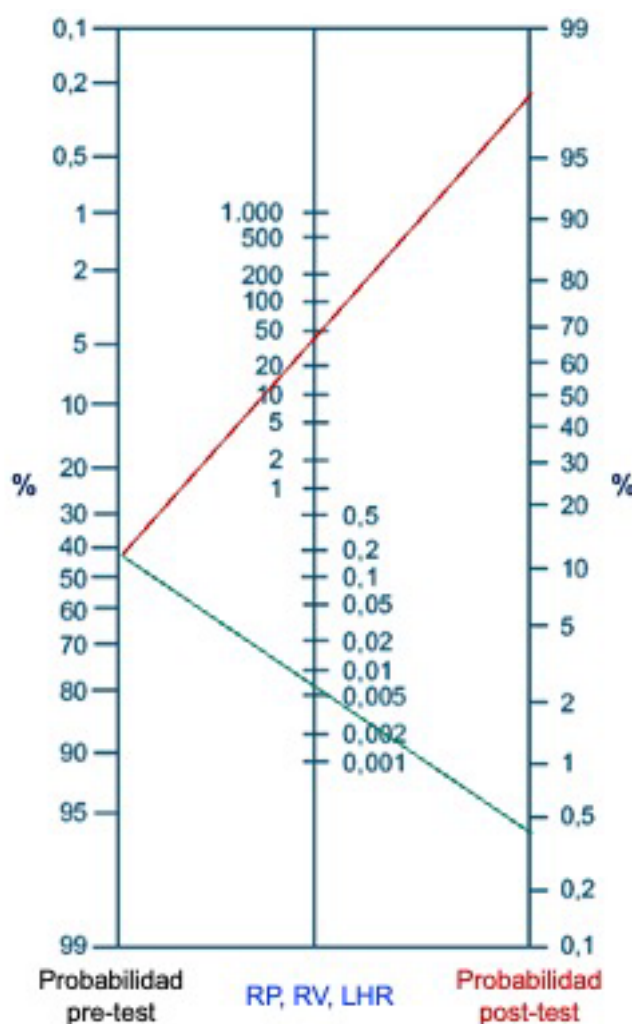


Fig. 4. Utilización del nomograma de Fagan. En un estudio de corte transversal, se verificó una prevalencia de enfermedad por reflujo gastroesofágico de 44,8 % en población general de una ciudad (Manterola *et al.*, 2020); lo que representa una probabilidad pre-test de que un sujeto de esa ciudad tenga el EI. Si al aplicar una PD con S: 91,6 %; E: 94,4 %; RV positiva: 86 % y RV negativa: 0,009 (Manterola *et al.*, 2002); en el mismo sujeto, si la PD es positiva para el EI, la probabilidad post-test de que el sujeto tenga el EI en estudio será del 97,0 %; y, si al aplicarla es negativa, la probabilidad post-test de que el sujeto tenga el EI en estudio será del 0,45 %.

Probabilidad post-test. Corresponde a la probabilidad que un sujeto tenga la enfermedad o EI dado un resultado positivo de la PD en estudio. Esto dependerá de: la prevalencia del EI, de la probabilidad pre-test (basada en el proceso diagnóstico), y del valor de las RV, RP o LHR.

Nomograma de Fagan. Fue descrito en 1975 por Fagan. Corresponde a una conversión a función sumatoria lineal simple (teorema Bayes). Está compuesto por tres columnas: probabilidad pre-test (prevalencia), RV o LHR, y la probabilidad pos-test. De tal modo que al trazar una línea entre la probabilidad pre-test y el LHR, la prolongación de la línea cortará en la probabilidad post-test. Se desarrolla un ejemplo de cómo la probabilidad pre-test (prevalencia), cambia la probabilidad pos-test al considerar las RV o LHR (Fig. 4).

EXACTITUD DE UNA PD

Es la capacidad de una PD de detectar correctamente la presencia o ausencia de un EI. Es decir, cuál la probabilidad de una PD prediga de forma correcta la positividad y negatividad de un EI o enfermedad en estudio (Fig. 1), lo que se calcula mediante la siguiente fórmula: Exactitud = $[(VP + VN) / \text{todos los sujetos en estudio}] \times 100$

Representa a la frecuencia de sujetos en los que la PD demuestra su condición de EI positivo o negativo. No obstante ello, es importante señalar que una PD puede ser más exacta para detectar positividad o negatividad del EI, por lo que puede ser ventajosa en diferentes escenarios.

Depende de la prevalencia del EI o enfermedad en estudio, por lo que no es útil para valorar comparativamente una PD respecto de otras, especialmente en resultados binarios o dicotómicos.

SENSIBILIDAD AL CAMBIO DE UNA PD

Corresponde a una medida de asociación entre el cambio en el puntaje observado y el cambio en el verdadero valor del constructo. Se aplica para medir diferencias en la cantidad de cambio entre individuos en estudio; para identificar factores que se asocian al cambio; y para inferir efectos de un tratamiento.

USO DE PD MÚLTIPLES

La indicación de aplicar de PDs múltiples es fundamentalmente cuando las PD existentes son imperfectas para el EI que se pretende diagnosticar; es decir, PD con S y E inferior a 80 %. En estas situaciones, se pueden utilizar PD en paralelo, o PD en serie:

- PD en paralelo: Aplica en situaciones de urgencia. En estos casos, se realizan todas las PDs necesarias, independiente de sus resultados. Así, cualquiera de ellas que resulte positiva es considerada como a favor del diagnóstico de enfermedad (+) o de EI positivo. Es una estrategia más sensible, pero que incrementa los FP.

Un ejemplo de la aplicación de esta estrategia se puede verificar en una revisión sistemática realizada con el objetivo de evaluar la precisión diagnóstica de PD y estrategias de cribado para tuberculosis en pacientes con VIH, utilizando como ER el cultivo de esputo (Dhana *et al.*, 2022). Utilizaron un panel de dos PD al mismo tiempo (W4SS y anomalías en la radiografía de tórax, en comparación con W4SS sola), y se constató que el panel incrementó la S (89 % [70 - 97] vs 53 % [35 - 71]) y disminuyó la E (33 % [17 - 54] vs 71 % [51 - 85]).

- PD en serie: Aplica en situaciones electivas o cuando alguna PD es muy costosa o riesgosa. En estos casos, las PDs se van aplicando de forma consecutiva, y el proceso se detiene al obtener un resultado negativo. Es una estrategia más específica, pero que incrementa los FN.

Un ejemplo de la aplicación de esta estrategia se puede verificar en un estudio de PD cuyo objetivo fue determinar características de tres PD para echinococcosis quística: inmunoprecipitación con DD5, ELISA-IgG (IgG) y electrotransferencia con Western Blot (WB). Una vez obtenidas sus respectivas S, E, VPP, VPN; se utilizaron dos estrategias de PD en serie distintas: IgG $\wedge$ DD5 y WB $\wedge$ DD5. Entonces, se verificó que con ambas estrategias se

incrementó discretamente la S (83,9 %), respecto de las tres PD por separado (69,3 - 82,7), así como el VPP (94,5 % para ambas estrategias, respecto de las PD aplicadas por separado, cuyas cifras fluctuaron entre 86,1 % y 92,9 %). Sin embargo, se constató un decremento en el comportamiento de la E y el VPN (Manterola *et al.*, 2005).

TAMIZAJE, CRIBADO O SCREENING

Se trata de un procedimiento diagnóstico indicado y utilizado para detección precoz de un EI o enfermedad, con el propósito de reducir secuelas; cuya aplicación permite obtener una estimación aproximada de la prevalencia o incidencia de un EI; así como de la carga de enfermedad; y por ende, de gran beneficio en vigilancia epidemiológica.

No obstante, para realizar tamizaje, se requiere un EI altamente prevalente en la población y PD con alta sensibilidad, de fácil aplicación, no invasivas, y bajo costo; aplicadas en poblaciones que aparentemente no tienen el EI en estudio (Fig. 5).

Permite seleccionar sujetos con resultado anormal de la PD, sin pretender un diagnóstico definitivo del EI; pero que pesquise al positivo como positivo y al negativo como negativo.

Las características de un tamizaje desde el punto de vista de una enfermedad o EI son la relevancia e interés por parte de la población general el poder diagnosticarlo precozmente; así como el disponer de mayor conocimiento respecto de la historia natural y epidemiología del EI (por

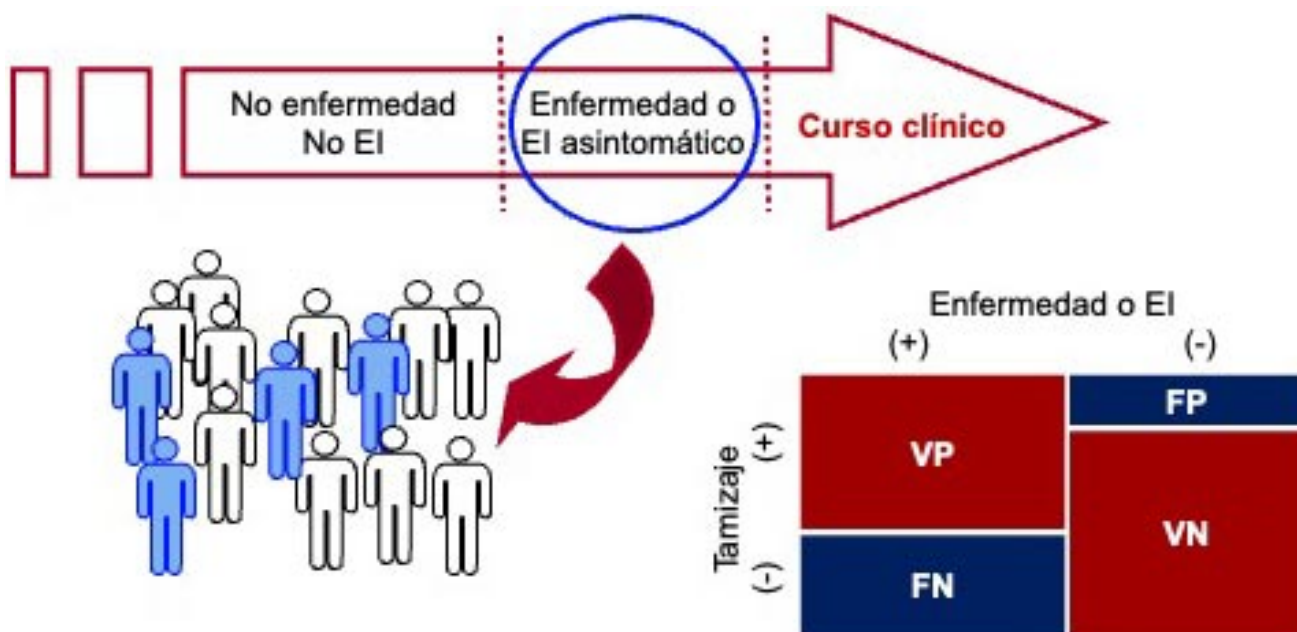


Fig. 5. Esquema del concepto de tamizaje y de su cálculo.

ejemplo, posibles factores de riesgo). Desde el punto de vista de la PD, esta debe contar con las siguientes propiedades: ser válida, precisa y simple de aplicar; con una distribución de valores conocida; un adecuado punto de corte que defina la positividad, y que sea aceptable para la población a la cual será aplicada. Desde la perspectiva de la eventual intervención que se aplicará en sujetos con tamizaje positivo, se ha de considerar que esta sea accesible, con evidencia concreta de las indicaciones de su uso, y que sea efectiva para resolver el problema y se encuentre disponible para su aplicación. Finalmente, desde el punto de vista de un programa de tamizaje, este debe estar apoyado en estudios con evidencia basada en ensayos clínicos con asignación aleatoria; que permitan una reducción significativa de variables como morbilidad o mortalidad; que su aplicación determine beneficios que superan potenciales daños; y que sea costo-efectivo.

SESGOS MÁS FRECUENTES EN UN ESTUDIO DE PD

Al evaluar un estudio sobre PD, es necesario reconocer si se aplicaron estrategias adecuadas en la población a estudio, y considera los riesgos de sesgos más habituales, que son: sesgo de medición, sesgo de selección y sesgo de confusión. Sin embargo, hay algunos subtipos de estos, que aplican específicamente a los estudios de PD.

- Sesgo de ER imperfecto: Este, depende de si los errores en el ER (de mala calidad), se relacionan o no con los errores en la PD en evaluación. Cuando las razones de error son diferentes y las clasificaciones erróneas por parte del ER son independientes de las clasificaciones erróneas por parte de la PD en evaluación, tanto la S como la E de la PD en evaluación serán erróneamente menores (Hall *et al.*, 2019).
- Sesgo de espectro: Este tipo de sesgo, se relaciona con el problema del espectro de enfermedad y no enfermedad (EI positivo vs. EI negativo), condición compleja dicotomizada, cuando posiblemente se trata de una variable categórica. Habitualmente se denomina a los sujetos con EI como “positivos”; sin embargo, algunos tienen una enfermedad leve o temprana, y otros tienen una enfermedad grave o avanzada; por ende, en lugar de denominarlos simplemente como “positivos”, debería existir algo así como una gama de “positivos” (por ejemplo: leve, moderado, grave). Lo mismo ocurre con los sujetos sin EI, a quienes se categorizan como “negativos”, cuando esta situación puede incluir perfectamente otras enfermedades o EI simultáneos y diferentes estados de salud (Hall *et al.*, 2019). Es decir, termina echándose en un mismo saco diferentes tipos de manzanas o de peras.

- Sesgo de incorporación: La positividad de la PD en evaluación no debe tener ningún rol en la determinación de si el ER clasifica al sujeto en estudio como positivo o no. Si la PD en evaluación se incorpora en el ER, tanto la S como la E estarán sesgadas al alza en comparación con un estudio que utiliza un ER independiente (Rendón-Macías *et al.*, 2020).

- Sesgo de verificación: También conocido como sesgo de procedimiento. Es un sesgo frecuente en este tipo de estudios. Ocurre cuando los sujetos en estudio que resultan positivos en la PD en evaluación tienen más probabilidades de recibir el ER, y solo aquellos que reciben el ER se incluyen en el estudio. La aplicación del ER no debe depender del resultado de la PD en evaluación (Kohn *et al.*, 2013).

- Sesgo de verificación diferencial: También conocido como sesgo del doble ER. Ocurre cuando se utiliza más de un ER, dependiendo de los resultados de la PD en evaluación; y también, cuando los dos ER pueden ofrecer respuestas diferentes (Rendón-Macías *et al.*, 2020).

NIVELES DE EVIDENCIA Y GRADOS DE RECOMENDACIÓN

La investigación en PD puede utilizar distintos diseños; desde descriptivos, en los que se comparan resultados de PD en pacientes con y sin el EI para mostrar que existe una diferencia en el resultado; pasando por estudios de corte transversal, en los que se emplea la PD en estudio y un ER en un único punto en el tiempo para determinar S, E, VP, etc.; hasta el desarrollo de revisiones sistemáticas de PD. Existen al menos tres propuestas de clasificación de la evidencia con relación a escenario de PD: la propuesta de Sackett (Canadá, 1996), la del National Health and Medical Research Council (NHMRC, Australia, 1999); y la del Centro de Medicina Basada en Evidencias de Oxford (Oxford Centre for Evidence-Based Medicine, 2009) (Manterola *et al.*, 2014; Wallace *et al.*, 2022) (Tabla II).

LISTAS DE VERIFICACIÓN PARA EL REPORTE DE RESULTADOS

Las opciones son pocas y dos de ellas muy específicas en su objetivo de uso.

1. **GRAS**. Guía de evaluación de la confiabilidad y acuerdo en mediciones en atención médica. Compuesta por 15 ítems agrupados en 6 dominios (Kottner *et al.*, 2011).
2. **STARD** (Standards for Reporting Diagnostic Accuracy). Su objetivo es generar un estándar para el informe de

Tabla II. Jerarquización de la evidencia para escenario de PD.

GR	NE	Sackett	CEBM, 2009	NHMRC
A	1a	RS de estudios de PD de nivel 1	RS de estudios de PD de alta calidad	I. RS de estudios nivel 1
	1b	Estudios de PD con ER de pacientes consecutivos	Estudios de cohortes que validen la calidad de una PD con un ER adecuado	II. Estudios de PD con comparación enmascarada e independiente y ER válido, entre sujetos consecutivos
	1c		Estudios de PD de alta S y E	
B	2a	RS de estudios de PD de nivel 2	RS de estudios de PD de nivel 2 con homogeneidad	III-1. Estudios de PD con comparación enmascarada e independiente y ER válido, entre sujetos no consecutivos
	2b	Estudios de PD con ER de pacientes no consecutivos	Estudios exploratorios que determinan factores significativos y validados con ER (por regresión logística)	III-2. Comparación con ER que no cumple con el criterio requerido para el NE II y III-1
	3a		RS de estudios 3b y mejor calidad	III-3. Estudios de C y C
	3b	Estudios sin ER, o no consecutivos	Estudios sin ER, o no consecutivos	
C	4	Estudios de C y C sin ER	Estudios de C y C sin ER	IV. Estudios de PD sin ER
D	5	Opinión de expertos	Opinión de expertos	

NE: niveles de evidencia. PD: prueba diagnóstica. CEBM: Oxford Centre for Evidence Based Medicine. ER: estándar de referencia. NHMRC: National Health and Medical Research Council. C y C: casos y controles. S: sensibilidad. E: especificidad.

estudios de precisión diagnóstica. Compuesto por 30 ítems agrupados en 7 dominios, un diagrama de flujo y asignación de puntaje (Bossuyt *et al.*, 2015). Es la más utilizada a lo largo del tiempo.

3. TRIPOD. Declaración para mejorar la transparencia en el reporte de un estudio de modelos de predicción para diagnóstico individual y pronóstico, independientemente del tipo de estudio utilizado. Consta de 6 dimensiones y 22 ítems (Collins *et al.*, 2015).

ESCALAS DE CALIDAD METODOLÓGICA

Las opciones son pocas y tres de ellas muy específicas en su objetivo de uso.

1. Arrivé: Instrumento construido con el objetivo de evaluar CM de las investigaciones clínicas que utilizan exámenes radiológicos. Compuesto por 15 criterios relacionados con el diseño de estudio y las características de la población. Cuenta con validez de fachada y evaluación de confiabilidad interobservador para estándares individuales, que se midió mediante grado de acuerdo porcentual y estadística k (0,74 a 0,97); y para el total se aplicó CCI, obteniéndose un $r=0,91$ (Arrivé *et al.*, 2000).

2. QUADAS-2. Herramienta para la evaluación de la calidad de estudios de precisión diagnóstica incluidos en una revisión sistemática (RS). Se aplica en 4 fases: resumen de la pregunta, adaptación al estudio analizado, diagrama de flujo para los estudios primarios y evaluación del riesgo de sesgo y aplicabilidad (Whiting *et al.*, 2011).

3. QAREL. Herramienta de evaluación de la confiabilidad de los estudios de PD, que también pudiera ser utilizada en RS de PD. Compuesta por 7 dominios y 11 ítems. Se aplica con base en preguntas de 3 alternativas de respuesta: "sí" (buena calidad), "no" (mala calidad), "no claro"; y algunos artículos incluyen la opción "no aplicable" (Lucas *et al.*, 2013).

4. MInCir-Dg (Metodología de Investigación en Cirugía Diagnóstico). Diseñada para evaluar CM de estudios de diagnóstico. Está compuesta por 9 ítems agrupados en 3 dominios. La puntuación final puede variar de 9 a 45 puntos, con un punto de corte de 24 puntos que define el constructo. Cuenta con evaluación de validez de fachada, contenido (80-100 % para todos los ítems incluidos) de criterio y constructo ($p<0,05$); y estudio de confiabilidad interobservador con un CCI de 0,96 (Manterola *et al.*, 2016).

MANTEROLA, C.; BENDEL, S.; RIVADENEIRA, J. & HOLGUÍN, J. P. Concepts regarding diagnostic tests and their application. *Int. J. Morphol.*, 43(5):1785-1795, 2025.

SUMMARY: With the continuous technological advancement, diagnostic tests (DT) keep improving, presenting us with new alternatives. Given the large number of DTs in use, it is necessary to periodically evaluate the various existing options in order to determine their role in clinical practice. First, the accuracy of the DT must be determined, including its sensitivity and specificity, predictive values, likelihood ratios, and receiver operating characteristic (ROC) curves. Then, its role in a diagnostic pathway must be determined, and finally, its effect on patient outcomes should be examined. In this manuscript, we will outline some principles regarding the reference standard, properties, and discriminatory power of a DT (validity and reliability of a diagnostic

test, likelihood ratios, pre- and post-test probabilities, ROC curves, common biases in diagnostic test studies, levels of evidence, and grades of recommendation in DT studies; application of checklists and quality assessment scales for DT studies; and concepts on the use of multiple DT and the application of screening strategies). The aim of this manuscript was to create a study document to review and discuss essential features of DT, providing some examples for better understanding.

KEY WORDS: "Diagnosis"[Mesh]; "Diagnostic Tests, Routine"[Mesh]; "Diagnostic Techniques and Procedures"[Mesh]; Diagnostic accuracy; Diagnostic test accuracy; Diagnostic test properties; Sensitivity; Specificity; Positive predictive value; Negative predictive value; Receiver operating characteristic curve; Likelihood ratio.

REFERENCIAS BIBLIOGRÁFICAS

- Arrivé, L.; Renard, R.; Carrat, F.; Belkacem, A.; Dahan, H.; Le Hir P.; Monnier-Cholley, L. & Tubiana, J. M. A scale of methodological quality for clinical studies of radiologic examinations. *Radiology*, 217(1):69-74, 2000.
- Berger, D. A brief history of medical diagnosis and the birth of the clinical laboratory. Part 1--Ancient times through the 19th century. *MLO Med. Lab. Obs.*, 31(7):28-30, 1999a.
- Berger, D. A brief history of medical diagnosis and the birth of the clinical laboratory. Part 2--Laboratory science and professional certification in the 20th century. *MLO Med. Lab. Obs.*, 31(8):32-4, 36, 38, 1999b.
- Bossuyt, P. M.; Reitsma, J. B.; Bruns, D. E.; Gatsonis, C. A.; Glasziou, P. P.; Irwig, L.; Lijmer, J. G.; Moher, D.; Rennie, D.; de Vet, H. C.; et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, 351:h5527, 2015.
- Burgos, D. M. E. & Manterola, D. C. Cómo interpretar un artículo sobre pruebas diagnósticas. *Rev. Chil. Cir.*, 62(3):301-8, 2010.
- Carter, J. V.; Pan, J.; Rai, S. N. & Galandiuk, S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*, 159(6):1638-45, 2016.
- Chassé, M. & Fergusson, D. A. Diagnostic accuracy studies. *Semin. Nucl. Med.*, 49(2):87-93, 2019.
- Collins, G. S.; Reitsma, J. B.; Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*, 350:g7594, 2015.
- Dhana, A.; Hamada, Y.; Kengne, A. P.; Kerkhoff, A. D.; Rangaka, M. X.; Kredo, T.; Baddeley, A.; Miller, C.; Singh, S.; Hanifa, Y.; et al. Tuberculosis screening among ambulatory people living with HIV: A systematic review and individual participant data meta-analysis. *Lancet Infect. Dis.*, 22(4):507-18, 2022.
- Elston, D. M. Likelihood ratios. *J. Am. Acad. Dermatol.*, 86(6):1229, 2022.
- Hall, M. K.; Kea, B. & Wang, R. Recognising bias in studies of diagnostic tests Part 1: Patient selection. *Emerg. Med. J.*, 36(7):431-4, 2019.
- Kohn, M. A.; Carpenter, C. R. & Newman, T. B. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad. Emerg. Med.*, 20(11):1194-206, 2013.
- Kottner, J.; Audigé, L.; Brorson, S.; Donner, A.; Gajewski, B. J.; Hróbjartsson, A.; Roberts, C.; Shoukri, M. & Streiner, D. L. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J. Clin. Epidemiol.*, 64(1):96-106, 2011.
- Lucas, N.; Macaskill, P.; Irwig, L.; Moran, R.; Rickards, L.; Turner, R. & Bogduk, N. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med. Res. Methodol.*, 13:111, 2013.
- Manterola, C.; Grande, L.; Bustos, L. & Otzen, T. Prevalence of gastroesophageal reflux disease: A population-based cross-sectional study in southern Chile. *Gastroenterol. Rep. (Oxf.)*, 8(4):286-92, 2020.
- Manterola, C.; Cartes-Velásquez, R. & Otzen, T. Instructions for using the MInCir scale to assess methodological quality in diagnostic accuracy studies. *Int. J. Morphol.*, 34(1):78-84, 2016.
- Manterola, C.; Asenjo-Lobos, C. & Otzen, T. Hierarchy of evidence: levels of evidence and grades of recommendation from current use. *Rev. Chilena Infectol.*, 31(6):705-18, 2014.
- Manterola, C. Cómo interpretar un artículo sobre pruebas diagnósticas. *Rev. Med. Clin. Condes.*, 20(5):708-17, 2009.
- Manterola, C.; Cuadra, A.; Muñoz, S.; Sanhueza, A.; Bustos, L.; Vial, M. & Fonseca, F. In a diagnostic test study the validity of three serodiagnostic test was compared in patients with liver echinococcosis. *J. Clin. Epidemiol.*, 58(4):401-6, 2005.
- Manterola, C.; Muñoz, S.; Grande, L. & Bustos, L. Initial validation of a questionnaire for detecting gastroesophageal reflux disease in epidemiological settings. *J. Clin. Epidemiol.*, 55(10):1041-5, 2002.
- Nahm, F. S. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J. Anesthesiol.*, 75(1):25-36, 2022.
- Ost, D. E. Interpretation and application of the likelihood ratio to clinical practice in thoracic oncology. *J. Bronchology Interv. Pulmonol.*, 29(1):62-70, 2022.
- Oxford Centre for Evidence-Based Medicine. *Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009)*. Oxford, Centre for Evidence-Based Medicine, University of Oxford; 2009. Disponible en: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009>
- Rendón-Macías, M. E.; Valenzuela, M. & Villasis-Keever, M. Á. Biases in diagnostic test studies: impact on estimating sensitivity and specificity. *Rev. Alerg. Mex.*, 67(2):165-73, 2020.
- Shreffler, J. & Huecker, M. R. *Type I and Type II Errors and Statistical Power*. In: StatPearls. Internet. Treasure Island (FL), StatPearls Publishing, 2024.
- Timsit, E.; Léguillette, R.; White, B. J.; Larson, R. L. & Buczinski, S. Likelihood ratios: an intuitive tool for incorporating diagnostic test results into decision-making. *J. Am. Vet. Med. Assoc.*, 252(11):1362-66, 2018.
- Torres-Quevedo, R.; Manterola, C.; Sanhueza, A.; Bustos, L.; Pineda, V. & Vial, M. Diagnostic properties of a symptoms scale for diagnosing reflux esophagitis. *J. Clin. Epidemiol.*, 62(1):97-101, 2009.
- Vali, Y.; Yang, B.; Olsen, M.; Leeftang, M. M. G. & Bossuyt, P. M. M. Reporting of test comparisons in diagnostic accuracy studies: a literature review. *Res. Synth. Methods*, 12(3):357-67, 2021.
- Wallace, S. S.; Barak, G.; Truong, G. & Parker, M. W. Hierarchy of evidence within the medical literature. *Hosp. Pediatr.*, 12(8):745-50, 2022.
- Walker, H. K. *Clinical Methods: The History, Physical, and Laboratory Examinations*. Chapter 1. En: Walker, H. K.; Hall, W. D. & Hurst, J. W. (Eds.). *The Origins of the History and Physical Examination*. 3rd ed. Boston, Butterworths, 1990.
- Whiting, P. F.; Rutjes, A. W.; Westwood, M. E.; Mallett, S.; Deeks, J. J.; Reitsma, J. B.; Leeftang, M. M.; Sterne, J. A.; Bossuyt, P. M. & QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.*, 155(8):529-36, 2011.

Correspondencia:

Dr. Carlos Manterola

CEMyQ

Universidad de La Frontera.

Francisco Salazar 01145

Temuco

CHILE

E-mail: carlos.manterola@ufrontera.cl