



Comentario: Sensibilidad, especificidad y valores predictivos: fundamentos, flexibilidades y dificultades en la investigación y la práctica

Gilat Grunau¹ y Shai Linn^{2*}

¹Departamento de Radiología, Hospital General de Vancouver, Universidad de Columbia Británica, Vancouver, BC, Canadá; ²Escuela de Salud Pública, Universidad de Haifa, Haifa, Israel

Palabras clave: teorema de Bayes, clínica, diagnóstico, epidemiología, métodos, screening.

Un comentario sobre

Sensibilidad, especificidad y valores predictivos: fundamentos, flexibilidades y dificultades en la investigación y la práctica

por Trevethan R. (2017). Frente. Salud Pública 5:307. doi: 10.3389/fpubh.2017.00307

Nos gustaría sugerir nuevas perspectivas, siguiendo el artículo de Trevethan (1), sobre conceptos erróneos sobre las medidas de detección.

MEDIDAS DE PRECISIÓN DE LAS PRUEBAS EN DOS SITUACIONES DISTINTAS: DETECCIÓN Y DIAGNÓSTICO (2–20)

Sugerimos una distinción clara entre las medidas de precisión de la **detección de una enfermedad en un entorno de detección en poblaciones**, utilizando datos sobre personas con y sin estado de enfermedad conocido, versus medidas de precisión de **diagnóstico en un entorno clínico** en individuos cuando se desconoce el estado de la enfermedad.

En la primera situación, un investigador está realizando un estudio en el que la prevalencia de la enfermedad es **artificial**: lo determina el investigador en función del número de personas con y sin la enfermedad que se incluyen en el estudio. Por ejemplo, si en un estudio se examina la sensibilidad y especificidad de una prueba en 100 personas con una enfermedad (por ejemplo, SIDA) y 100 personas sin la enfermedad, la prevalencia del SIDA en este estudio en particular es del 50%, lo que por supuesto está lejos de la verdadera prevalencia del SIDA. El sensibilidad y especificidad se utilizan para describir el técnico características de una prueba, es decir, cuántas personas con o sin enfermedad serán **detectado correctamente** mediante una prueba en una población con diagnósticos conocidos. Estas medidas son importantes en la salud pública y la planificación sanitaria. Por ejemplo, uno podría necesitar saber el porcentaje de personas enfermas (sensibilidad) o personas sanas (las especificidad) que se detectarán entre los viajeros en un aeropuerto y así planificar medidas preventivas en tiempos de epidemia de enfermedades transmisibles.

La segunda situación es un entorno clínico en el que aún se desconocen los diagnósticos y la prueba se utiliza para diagnosticar la enfermedad en individuos: el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN) son una estimación de la precisión de la prueba, es decir, de las fracciones de pacientes que son **diagnosticado correctamente** como positivo o negativo, respectivamente.

Expliquemos la necesidad de utilizar dos enfoques diferentes para las dos situaciones distintas descritas anteriormente, con notaciones análogas específicas (cursiva, minúscula y mayúscula) y las ecuaciones derivadas.

Estimación de la precisión de la detectabilidad de las pruebas en poblaciones (Tabla 1)

Usamos cursiva **letras minúsculas** en la descripción del cribado en la población general en un 2×2 mesa, **tabla 1**. El sensibilidad y especificidad se calculan en muestras de personas con (a + c) y sin (b + d) la enfermedad en una población seleccionada. En esta tabla, no es apropiado incluir totales del eje “horizontal” de los resultados de la prueba (T).

OPEN ACCESS

Editado por:

Aarón A. Alford,
Red Nacional de Salud Pública
Institutos, Estados Unidos

Revisado por:

James Harris Derzón,
RTI Internacional, Estados Unidos

* Correspondencia:

Shai Linn
slinn@univ.haifa.ac.il

Sección de especialidad:

Este artículo fue enviado a
Epidemiología,
una sección de la revista
Fronteras en Salud Pública

Recibió: 24 mayo 2018

Aceptado: 17 agosto 2018

Publicado: 28 de septiembre de 2018

Citación:

Comentario de Grunau G y Linn S
(2018): Sensibilidad, especificidad y valores
predictivos: fundamentos, flexibilidades y
dificultades en la investigación y
práctica. Frente. Salud Pública 6:256.
doi: 10.3389/fpubh.2018.00256

La razón de probabilidad positiva, por favor, es la relación de sensibilidad hacia la fracción de falsos positivos, *fpf* (es decir, 1-especificidad). Obviamente, cuando $por\ favor = 1$, la prueba es inútil: su detección será igualmente correcta entre personas con la enfermedad y personas sin la enfermedad, ($sensibilidad = fpf$). Cuando $por\ favor > 1$, cuanto mayor sea el valor de *por favor*, cuanto más eficaz sea la prueba para detectar correctamente a las personas con la enfermedad, será menos probable que identifique erróneamente a una persona sana como una persona con la enfermedad. Cuando $por\ favor < 1$, las sensibilidades son menores que el *fpf*, para que la prueba detecte correctamente a las personas con la enfermedad con menos frecuencia que la que identifica incorrectamente a una persona sana como una persona con una enfermedad; es decir, la prueba es más engañosa que útil para detectar una enfermedad.

De manera similar, el índice de probabilidad negativo, *nlr*, es la relación de la fracción de falso negativo, *fnf* (es decir, 1-sensibilidad) a la especificidad. Cuando $nlr = 1$, la prueba es inútil: tampoco detectará la enfermedad incorrectamente entre personas con la enfermedad (y no detectará la enfermedad) y sin la enfermedad (especificidad = *fnf*).

Cuando $nlr < 1$, la prueba es más eficaz: identifica correctamente a las personas sanas con más frecuencia que a las personas que padecen una enfermedad como si no la padecen. Cuando $nlr > 1$, la prueba es más engañosa que útil para detectar la ausencia de una enfermedad.

Estimación de la precisión del diagnóstico de una enfermedad en la población de pacientes (Tabla 2)

La aplicación de una prueba de diagnóstico a una población de pacientes (objetivo) utiliza un método similar 2x2 mesa (Tabla 2). Para evaluar la eficacia de la aplicación de una prueba de diagnóstico en la población de pacientes, el investigador primero observa el resultado, es decir, los resultados de la prueba, y obtiene información sobre el factor del estudio, es decir, el estado de la enfermedad.

Usamos **letras mayúsculas** para describir el cribado en la población de pacientes en un 2x2 mesa, **Tabla 2**. Son los datos de esta tabla.

que sean de interés para el paciente (y el médico), respondiendo las siguientes preguntas: Cuando la prueba es positiva, ¿cuál es la probabilidad de que el paciente tenga la enfermedad? (Responsable por el PPV); y cuando la prueba es negativa ¿cuál es la probabilidad de que el paciente no tenga la enfermedad? (Responsable por el VPN).

Hemos sugerido índices de confiabilidad de una prueba, análogos a los índices de probabilidad que se han discutido anteriormente (19).

El índice predictivo positivo (PPR) es la relación entre el VPP y la fracción falsa negativa en la población de pacientes (FNF). Obviamente, cuando $PPR = 1$ la prueba es inútil: diagnosticará correctamente (cuando la prueba es positiva) y no diagnosticará a personas con la enfermedad ($PPV = FNF$).

Por el contrario, cuando $PPR > 1$, cuanto mayor sea el valor de PPR, más eficaz será la prueba para diagnosticar correctamente a las personas con la enfermedad en relación con el diagnóstico de la ausencia de la enfermedad cuando una persona sí la tiene, por lo que no se puede obtener un resultado positivo. diagnóstico. Cuando $PPR < 1$, la prueba es engañosa. Diagnostica a personas enfermas como si no tuvieran una enfermedad con más frecuencia que a una persona con una enfermedad.

De manera similar, el índice de predicción negativa (NPR) es la relación entre la fracción de falsos positivos en la población de pacientes (FPF) y el VPN. Cuando $NPR = 1$, la prueba es inútil: diagnosticará correctamente a personas sin la enfermedad (cuando la prueba es negativa) y no diagnosticará a personas sin la enfermedad ($NPV = FPF$). Por el contrario, cuando $NPR < 1$, cuanto menor es el valor de NPR, más eficaz es la prueba, ya que diagnostica correctamente a personas sanas como personas sin una enfermedad con más frecuencia que diagnostica incorrectamente que una persona sin enfermedad tiene la enfermedad. Cuando $NPR > 1$, la prueba es engañosa. Diagnostica a personas sanas con una enfermedad con más frecuencia que a personas sanas.

Tabla 1 | Presentación de datos en una población seleccionada, evaluando la detectabilidad de una prueba.

		Estado de la enfermedad	
		S+	S-
Resultados de la prueba	T+	a	b
	-T	c	d
		a+c	b+d

$$sensibilidad = PAG(t+ | S+) = \frac{a}{a+c}, especificidad = PAG(t- | S-) = \frac{d}{b+d}$$

$$Ratio\ de\ probabilidad\ positivo = plr = \frac{sensibilidad}{1-especificidad}$$

$$Razón\ de\ probabilidad\ negativa = nlr = 1 - \frac{sensibilidad}{especificidad}$$

Términos de error para la población de

$$estudio: Fracción\ falsa\ positiva = fpf = \frac{b}{b+d}$$

$$Fracción\ falsa\ negativa = fnf = \frac{c}{a+c}$$

Tabla 2 | Presentación de datos en el entorno de un estudio clínico en una población de pacientes objetivo que evalúa la capacidad de diagnóstico de una prueba.

		Estado de la enfermedad		
		S+	S-	
Resultados de la prueba	T+	A	B	A+B
	-T	C	D	C+D

$$PPV = PAG(S+ | t+) = \frac{A}{A+B}, VPN = PAG(S- | t-) = \frac{D}{C+D}$$

$$Ratio\ predictivo\ positivo = PPR = \frac{PPV}{1-VAN}$$

$$Ratio\ predictivo\ negativo = NPR = 1 - \frac{PPV}{VPN}$$

error para la población de pacientes:

$$Fracción\ de\ falsos\ positivos = FPF = \frac{B}{A+B} = 1 - VPP$$

$$Fracción\ falsa\ negativa = FNF = \frac{C}{C+D} = 1 - NPV$$

GANANCIA EN CERTEZA Y MEDIDAS RESUMEN

Índice de Youden como medida resumida de la detectabilidad de una prueba en poblaciones

En 1950, Youden propuso un índice como medida de la bondad de una prueba. Usando las fracciones de falsos positivos y falsos negativos, el índice se define como

$$j = 1 - (fpr + fnr) \\ = (1 - fpr) + (1 - fnr) - 1 = \text{sensibilidad} + \text{especificidad} - 1$$

Cuando $j = 1$, la prueba siempre es correcta. En otras palabras, no hay errores, por lo que $fpr + fnr = 0$; es decir, la prueba detecta correctamente el estado de enfermedad. Cuando $j = 0$, suponiendo que sensibilidad y especificidad tienen igual importancia en la determinación de la ganancia esperada, la ecuación anterior implica que cuando $\text{sensibilidad} + \text{especificidad} = 1$, la prueba no proporciona información general. En otras palabras, la prueba es inútil si la proporción de errores es igual al 100%, es decir, cuando $fpr + fnr = 1$, $j = 0$. Cuando $j < 0$ (entre -1 y 0), la prueba es engañosa; es decir, los resultados de las pruebas se asocian negativamente con el verdadero diagnóstico.

también se puede interpretar como la probabilidad obtenida de información de detección correcta, es decir, la diferencia entre las probabilidades conjuntas de detección correcta (detección positiva o negativa), $\text{sensibilidad} \times \text{especificidad}$, vs. las probabilidades conjuntas de detección incorrecta, $fpr \times fnr$.

$$J = \text{sensibilidad} \times \text{especificidad} - (fpr \times fnr) \\ = \text{sensibilidad} \times \text{especificidad} - 1 - (\text{sensibilidad} \times 1 - \text{especificidad}) = \\ \text{sensibilidad} + \text{especificidad} - 1$$

Índice resumido predictivo como medida resumida de la diagnosticabilidad de una prueba en individuos (20)

Hemos propuesto un índice resumen de la información, el Índice Resumen Predictivo (PSI = 9) (17, Tabla 4), que es una medida de la información adicional proporcionada por los resultados de la prueba, más allá del conocimiento previo (la prevalencia de la enfermedad). Así, el PSI resume la información en **Tabla 2**. Tengamos en cuenta que la información de un resultado positivo de la prueba más allá de lo que ya se sabe a priori sobre la prevalencia de la enfermedad es PPV-Prevalencia. De manera similar, la información de un resultado negativo de la prueba más allá de lo que ya se sabe a priori sobre la probabilidad de que no haya enfermedad (la prevalencia de ninguna enfermedad) es $1 - \text{VPN}$ (1-Prevalencia).

Así, la información global, es decir, la ganancia en certeza que obtenemos después de realizar una prueba, más allá de lo que ya se sabe, es calculable como una medida resumida.

$$\text{Ganancia total en certeza} = \text{VPP} - \text{Prevalencia} \\ + \text{VPN} - (1 - \text{Prevalencia}) \\ = \text{VPP} + \text{VPN} - 1 = 9$$

Alternativamente, es la información que no se deriva de errores, FNF y FPF

$$9 = 1 - (\text{FPF} + \text{FNF}) = 1 - [(1 - \text{VPP}) + (1 - \text{VPN})] \\ = \text{VPP} + \text{VPN} - 1$$

Si $9 = 1$, la prueba siempre es correcta, es decir, no hay errores, por lo que $\text{FPF} + \text{FNF} = 0$; es decir, la prueba detecta correctamente el estado de enfermedad. Cuando $9 = 0$, $\text{VPP} + \text{VPN} = 1$ y la prueba no proporciona información general. En otras palabras, la prueba es inútil si la proporción de errores es igual al 100%, es decir, cuando $\text{FPF} + \text{FNF} = 1$, $9 = 0$. Por ejemplo, si los resultados de la prueba son aleatorios y la probabilidad de FPF y FNF es del 50%, entonces la prueba es inútil. Cuando $9 < 0$, valores negativos (entre -1 y 0) de 9 hacen que la prueba sea engañosa; es decir, los resultados de las pruebas se asocian negativamente con el verdadero diagnóstico.

El PSI también se puede interpretar como la probabilidad obtenida de información de diagnóstico correcto, es decir, la diferencia entre las probabilidades conjuntas de diagnóstico correcto (diagnóstico positivo o negativo) $\text{PPV} \times \text{VPN}$ vs. probabilidades conjuntas de diagnóstico incorrecto $\text{FPF} \times \text{FNF}$

$$\text{PPV} \times \text{VPN} - \text{FNR} \times \text{FPR} \\ = \text{PPV} \times \text{VPN} - [(1 - \text{VPP}) \times (1 - \text{VPN})] \\ = \text{PPV} \times \text{VPN} - 1 + \text{VPN} + \text{VPP} - \text{VPN} \times \text{VPP} = \psi$$

UTILIZACIÓN DEL TEOREMA DE BAYES (1-20)

Hemos discutido el uso de dos enfoques diferentes para la detección versus el diagnóstico. En la práctica, los análisis son frecuentemente realiza en dos etapas: la primera utiliza un selectivo (estudio) población para la cual sensibilidad y especificidad se calculan, comparando una prueba de detección basada en una prueba que es menos precisa, costosa o invasiva con un estándar de oro (GS), que puede ser una prueba más precisa pero más costosa o invasiva (p. ej., biopsia), y luego, el método de Bayes. Se utilizan el teorema y la prevalencia, junto con el sensibilidad y especificidad , para calcular el VPP o el VPN . Esto es particularmente útil cuando no tenemos la información necesaria para construir **Tabla 2**; es decir, cuando no podemos calcular directamente el VPP y el VPN , porque con frecuencia es inviable y poco ético realizar tanto las pruebas de diagnóstico como una prueba definitiva adicional (el "estándar de oro" contra el cual se evalúa la prueba) para determinar el diagnóstico verdadero. Por lo tanto, la prueba se evalúa en una muestra seleccionada de una población y sólo **tabla 1** se puede construir. Luego, el VPP y el VPN se calculan a partir de sensibilidad y especificidad y la prevalencia de la enfermedad, aplicando el teorema de Bayes.

DISCUSIÓN

El artículo de Trevethan explica la diferencia entre una "prueba de detección" y una prueba de diagnóstico. Agregamos y sugerimos que se debe hacer una distinción entre **detección** (en una población ya diagnosticada) y **diagnóstico** (en la población objetivo de pacientes).

El término "cribado" se utiliza en dos situaciones diferentes. El cribado se puede realizar en un **población** para detectar una salud

condición para fines de salud pública. Luego, esta actividad se evalúa en estudios entre personas con un estado de enfermedad conocido, calculando el VPP y el FPF entre personas con una enfermedad ya diagnosticada y entre aquellas sin enfermedad, calculando la especificidad y el FPF . La prevalencia de la enfermedad en tales estudios es frecuentemente artificial: la selección de participantes con o sin enfermedad puede depender del presupuesto, la disponibilidad de personas con un estado de enfermedad conocido (con o sin enfermedad) y consideraciones éticas. Por lo tanto, **Los valores predictivos no se pueden calcular en estos estudios.** Dichos estudios pueden ayudar en la planificación de los servicios de salud en un entorno de salud pública y el posible uso de la prueba con fines de salud pública (rentabilidad y utilidad).

Por el contrario, la detección de una enfermedad en la población de pacientes en un entorno clínico se realiza con fines de diagnóstico y tratamiento. Esta actividad luego se evalúa en estudios entre personas que se presentan en las clínicas, sin un estado de enfermedad conocido. De hecho, la interacción clínica suele centrarse en la determinación del estado de una enfermedad. Por lo tanto, la detección en el entorno clínico se realiza para **diagnóstico** propósitos. En el ámbito clínico, es posible calcular los valores predictivos de una prueba, calculando el VPP y FPF entre personas con un resultado positivo y el VPN y FNF entre personas con resultados negativos. La prevalencia de la enfermedad en estudios en el ámbito clínico es frecuentemente la prevalencia real en la población atendida por una clínica.

Además, sugerimos que se preste atención a las medidas de ratios que se basan en la sensibilidad y especificidad, es decir, los ratios de verosimilitud, PLRN . Estas medidas pueden facilitar la evaluación de una prueba en el ámbito de la salud pública. De manera similar, sugerimos que se preste atención a las medidas de ratios que se basan en el VPP y el VPN , es decir, el PPR y el NPR mencionados anteriormente. Estas medidas pueden facilitar la evaluación de una prueba en el entorno clínico e informar a pacientes y médicos sobre la confiabilidad de una prueba diagnóstica para diagnosticar una enfermedad (la PPR) o un estado de salud (la NPR).

Sugerimos que se preste atención a las medidas sumarias, que indican la ganancia en certeza en la **detección** de una enfermedad en un entorno de salud pública y **diagnóstico** de una enfermedad en la población de pacientes en el entorno clínico, utilizando medidas como el índice de Youden (j) y el PSI (ψ) (respectivamente).

Esperamos que nuestro enfoque aclare la metodología y el uso de medidas de detección en un entorno de detección versus medidas de diagnóstico en un entorno clínico.

CONTRIBUCIONES DE AUTOR

SL desarrolló las expresiones algebraicas y la teoría. GG desarrolló los conceptos, el manuscrito y revisó la metodología, el artículo y las conclusiones.

REFERENCIAS

- Trevethan R. Sensibilidad, especificidad y valores predictivos: fundamentos, flexibilidades y dificultades en la investigación y la práctica. *Front. Salud Pública* (2017) 5:307. doi: 10.3389/fpubh.2017.00307
- Riffenburgh RH. *Estadística en Medicina*. San Diego, CA: Prensa académica (1999).
- Hirsch RP, Riegelman RK. *Operaciones estadísticas: análisis de datos de investigaciones en salud*. Oxford: Ciencia de Blackwell (1996).
- Feinstein AR. *Principios de estadística médica*. Boca Ratón, FL: Chapman & Hall/CRC Press (2002).
- Weinstein MC, Finberg HV. *Análisis de decisión clínica*. Filadelfia, Pensilvania: W. B. Saunders (1980).
- Altman DG. *Estadísticas prácticas para la investigación médica*. Londres: Chapman & Hall (1991).
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Epidemiología clínica*. 2da ed. Boston, MA: Pequeño marrón (1991).
- Kraemer HC. *Evaluación de Pruebas Médicas: Pautas Objetivas y Cuantitativas*. Londres: Publicaciones Sage (1992).
- Weiss NS. *Epidemiología clínica: el estudio del resultado de la enfermedad*. Oxford: Oxford University Press (1996).
- Riegelman RK. *Estudiar un estudio y realizar una prueba: cómo leer la evidencia médica*. 4ª edición. Filadelfia, PA: Lippincott, Williams y Wilkins (2000).
- Knottnerus JA, van Weel C. Introducción general: evaluación de procedimientos de diagnóstico. En: Knottnerus JA, editor. *La base de evidencia del diagnóstico clínico*. Londres: BMJ Books (2002). pag. 1–18.
- Sackett D, Haynes RB. La arquitectura de la investigación diagnóstica. En: Knottnerus JA, editor. *La base de evidencia del diagnóstico clínico*. Londres: BMJ Books (2002). pag. 19–38.
- Zhou XH, Obuchowski NA, McClish DK. *Métodos estadísticos en medicina diagnóstica*. Nueva York, Nueva York: Wiley-Interscience (2002).
- Pepe MS. *La Evaluación Estadística de Pruebas Médicas para Clasificación y Predicción*. Serie de ciencia estadística de Oxford 28. Oxford: Oxford University Press (2003).
- Morabia A. (ed.). *La historia de los métodos y conceptos epidemiológicos*. Berlín: Birkhauser Verlag (2004).
- Rothman KJ, Lash YL, Groenlandia S. *Epidemiología moderna*. 3ª edición. Nueva York, Nueva York: Lippincott Williams & Wilkins (2008).
- Grobee DR, Hoes AW. *Epidemiología clínica*. Burlington, MA: Jones y Bartlett Learning (2015).
- Linn S. Un nuevo enfoque conceptual para enseñar la interpretación de pruebas clínicas. *J Educación Estadística*. (2004) 12. doi: 10.1080/10691898.2004.11910632
- Linn S. Las nuevas características de las pruebas de diagnóstico orientadas al paciente, análogas a los índices de probabilidad, transmiten información sobre la confiabilidad. *J Clin Epidemiol*. (2005) 58:450–7. doi: 10.1016/j.jclinepi.2004.07.009
- Linn S, Grunau DP. Nueva medida resumida orientada al paciente de la ganancia total neta en certeza para pruebas de diagnóstico dicotómicas. *Epidemiol Perspect Innovat*. (2006) 3:11. doi: 10.1186/1742-5573-3-11

Declaración de conflicto de intereses: Los autores declaran que la investigación se realizó en ausencia de relaciones comerciales o financieras que pudieran interpretarse como un potencial conflicto de intereses.

Copyright © 2018 Grunau y Linn. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia de Atribución Creative Commons (CC BY). Se permite el uso, distribución o reproducción en otros foros, siempre que se cite al autor(es) original(es) y al propietario(s) de los derechos de autor y se cite la publicación original en esta revista, de acuerdo con la práctica académica aceptada. No se permite ningún uso, distribución o reproducción que no cumpla con estos términos.