

Notas metodológicas

De la evidencia científica a las decisiones clínicas

Cuando pedir una prueba diagnóstica

5/12/2005

Martín Caicoya Gómez-Morán

Jefe de Servicio de Prevención de Riesgos Laborales del Principado de Asturias. Hospital Monte Naranco.

El radiólogo Z está examinando los estudios realizados ese día. Tiene en sus manos el solicitado por el doctor X, lee la petición y sonrío: "*siempre acierta, es un clínico excelente.*" Efectivamente, el paciente examinado tenía la alteración que se buscaba. El radiólogo piensa que si estuviera enfermo acudiría a él.

Ahora tiene en sus manos el estudio solicitado por el doctor Y, "*aquí la moneda al aire, piensa, unas veces acierta y otras no.*" Esta vez no tenía nada, "*no sé para qué lo pide.*"

Supongamos que ambos profesionales tienen la misma habilidad clínica y que han examinado a dos pacientes. Han coincidido en el grado de certeza sobre las hipótesis: el paciente A tiene un 90% de probabilidades de tener la enfermedad y el B un 50%. El doctor Y decide pedir la prueba sólo al paciente B.

Supongamos que la prueba radiológica que examina el radiólogo Z es positiva en el 95% de los enfermos, sensibilidad 95%, y negativa en el 95% de los sanos, especificidad 95%. Le sugiero que revise en la tabla 1 el valor predictivo positivo (VPP) y negativo (VPN) bajo la hipótesis de una probabilidad de enfermedad del 90% y del 50%. En el primer caso un resultado positivo apenas añade información y un resultado negativo introduce incertidumbre. En el segundo caso, con probabilidad de enfermar del 50%, la radiología aportó mucha información.

Tabla1: Valores predictivos y sensibilidad

P	S	E	VPP	VPN
0.90	0.95	0.95	0.99	0.65
0.50	0.95	0.95	0.95	0.95

$$VPP = \frac{P \times S}{P \times S + (1-E) \times (1-P)}$$

$$VPN = \frac{E \times (1-P)}{E \times (1-P) + (1-S) \times P}$$

La enseñanza de esto es que el buen clínico es el que pide pruebas cuando tiene dudas. En el proceso diagnóstico, realiza una comparación de la información obtenida sobre el paciente, con otros que conoce por su experiencia o estudios, asignándole una probabilidad de pertenecer a un grupo. El mejor rendimiento de una prueba diagnóstica es cuando el clínico piensa que el paciente examinado tiene una probabilidad entre el 40 y el 60% de tener la enfermedad.

El radiólogo Z estaba equivocado: El doctor Y es más prudente tomando decisiones, utiliza mejor los recursos y facilita con su actuación que el sistema pueda operar mejor. Porque si el doctor X pide la prueba "para quedarnos tranquilos" esa tranquilidad supone gastos prescindibles y demoras en la atención a otros pacientes con las potenciales repercusiones para su salud.

Es verdad que el médico debe ofrecerle a su paciente todo lo que considere necesario para su curación. Pero a la vez es un administrador de recursos, si da mucho a uno, otros se pueden quedar sin lo necesario. Y entre los recursos hay que contar el tiempo, la atención y esfuerzo.

Hay que matizar la afirmación de que cuando se tiene una considerable certeza de que el paciente tiene o no la enfermedad las pruebas complementarias son prescindibles. En realidad, la pregunta es: ¿qué beneficios obtendré de realizar más estudios y a qué coste? Pero tan importante como lo anterior, es decidir a qué nivel o grado de certeza de alteración se define positividad de la prueba. Ambas preguntas se pueden tratar con el análisis ROC y el análisis de decisión, técnicas que se basan en lo mismo.

Como definir positividad

A. La curva ROC

La curva ROC (Receiver Operator Characteristics) es una estrategia empleada por los radares para establecer el equilibrio entre detectar los casos, sensibilidad, y no tener muchos falsos positivos. Gráficamente se hace colocando los valores de 1-especificidad en el eje de las X y los correspondientes de

sensibilidad en el eje de las Y. Los puntos de cruce forman la curva ROC y la razón de verosimilitud (LR) es el cociente entre sensibilidad y frecuencia de falsos positivos (1-E), es decir la tangente a cada uno de los puntos de la curva ROC.

Supongamos una prueba numérica para la que conocemos la sensibilidad y especificidad en 4 niveles, tabla 2. El nivel 4 es el que más certeza diagnóstica ofrece: los así diagnosticados tienen 50 veces más probabilidades de ser enfermos que sanos. En cambio el nivel 1 no es informativo y los 2 y 3 los más equilibrados. Pero, ¿estaría satisfecho con el criterio 4? Evidentemente no porque deja fuera al 50% de los enfermos. Hay algo más que la certeza diagnóstica en esa decisión. La LR para el nivel 3, por cada unidad que se incrementa en el eje de las X (1-especificidad) se debe aumentar 3 en el de las Y (sensibilidad). ¿Cuál es, entonces, la mejor LR para mis propósitos o que punto de la curva ROC elijo?

Tabla 2: Razones de verosimilitud (LR) para distintos niveles de sensibilidad (S) y especificidad (E)

NIVEL	mg/dl	S	E	1 - E	LR
1	70	0.99	0.10	0,90	1.1
2	100	0.90	0.70	0,30	3
3	120	0.70	0.90	0,10	7
4	150	0.50	0.99	0,01	50

Fuente: *Diabetes Program Guide, Public Health Service 1960*

B. El análisis de decisión

El análisis de decisión es una técnica relativamente simple en la que se trata de comparar alternativas ante una decisión, asignando a cada una de ellas probabilidades y utilidades. Supongamos el caso del paciente con 90% de probabilidades de estar enfermo, ¿qué pasa si tomo la decisión de no hacer la radiología y lo trato como enfermo no estándolo? Imaginemos que lo que se dirime es si tiene un cáncer, por ejemplo de mama, se habrá realizado un tratamiento agresivo, posiblemente mutilante, además de calificar al paciente como portador de una enfermedad que va a condicionar su vida. En estos casos es evidente que la certeza diagnóstica ha de ser máxima. Porque el coste de tratar a un sano supera con mucho el de realizar las pruebas. Coste material y humano.

La expectativa de vida de un verdadero negativo (UVN), al que no se le hace nada será, evidentemente, del 100% del de su edad, sexo y procedencia geográfica, si no contamos para simplificar las molestias y efectos adversos de la prueba. Aceptamos en el supuesto anterior, que un enfermo tratado tiene una expectativa de vida ajustada por calidad del 85%, utilidad del verdadero positivo (UVP)=0.85*UVN y que el enfermo no tratado vive el 50%, utilidad del falso negativo (UFN)=0.5*UVN, y supongamos que tratar a un sano incurre en un perjuicio del 7% en su expectativa de vida, utilidad del falso positivo (UFP)=0.93*UVN.

Para cada enfermedad habrá un juego de perjuicios por no tratar al enfermo y tratar al sano que dependerá de la utilidad del tratamiento y de sus efectos secundarios. Por ejemplo, ante una meningitis meningocócica, los perjuicios de no tratar al enfermo son muy altos y los de tratar al sano son relativamente bajos. Pero ante un cáncer de páncreas los perjuicios de no tratar al enfermo no son tan espectaculares, pues la letalidad sigue siendo alta entre los tratados, y el perjuicio de tratar a un sano es bastante alto. A las utilidades hay que añadir el coste de tratar. De manera que en cada ejercicio se pueden variar la sensibilidad, la especificidad y las utilidades, así como la probabilidad de enfermedad.

Se hace normalmente en la clínica: se considera positiva una prueba de tuberculina con una induración mínima, 5 mm en lugares de alta prevalencia y se emplea un punto de corte más alto, 10 incluso 15 mm en lugares de baja prevalencia. Lo mismo se hace con la prueba de esfuerzo: descensos mínimos se consideran positivos en pacientes de alto riesgos y negativos en los de bajo riesgo. En realidad es como si se estuvieran aplicando dos pruebas diferentes.

Algebraicamente el rendimiento del test es la suma de las cuatro utilidades por la probabilidad de que ellas ocurran. Los cálculos de las diferentes utilidades se presentan en la tabla 3. En los enfermos (pE) se sumarán las utilidades de tratar a los enfermos que encuentre la prueba, que es la sensibilidad multiplicado por la UVP y las de no tratar a un enfermo, (1-sensibilidad) multiplicado por la UFP (1). En los casos en que la prueba se aplique a los que no tienen enfermedad (1-pE) expresión descrita en (2). La suma de ambos será la utilidad (U) de esa prueba (3). Por otra parte, la utilidad de tratar (UT) será la suma de la utilidad de tratar al sano y la de la de tratar el enfermo(4), la de no tratar (UNT) será (5).

Tabla 3: Expresión de las diferentes utilidades

(1)	UE	$pE \cdot \{S \cdot UVP + (1-S) \cdot UFN\}$
(2)	UNE	$(1-pE) \cdot \{(1-E) \cdot UFP + E \cdot UVN\}$
(3)	U	$pE \cdot \{S \cdot UVP + (1-S) \cdot UFN\} + (1-pE) \cdot \{(1-E) \cdot UFP + E \cdot UVN\}$
(4)	UT	$pE \cdot S \cdot UVP + (1-pE) \cdot (1-E) \cdot UFP$
(5)	UNT	$pE \cdot (1-S) \cdot UFN + (1-pE) \cdot E \cdot UVN$
(6)	Punto de corte	$\frac{UVN - UFP}{UVP - UFN} = \frac{1-pE}{pE}$
(7)	Punto de corte	$\frac{\text{Coste} = 1-pE}{\text{Beneficio} = pE}$

Sé que esto resulta muy antipático. Hay programas de análisis de decisión, pero a este nivel lo puede usted manejar con facilidad en una hoja de cálculo. Simplemente coloque en diferentes celdas los valores de sensibilidad, especificidad y UVP y UVN y haga que la celda que dan valor UFN sea igual a la que figura en UVP multiplicada por la fracción que haya decidido que sea la curación, lo mismo con UFN y los complementarios de sensibilidad y especificidad, la prevalencia y su complementario (tabla 4). Tendrá 10 celdas. En otras dos celdas calcule las utilidades en enfermedad y no-enfermedad en función de los valores, ver fórmulas (1) y (2) y súmelas en una nueva celda. Basta variar los valores de sensibilidad y especificidad para conocer a qué nivel de validez la prueba rinde más con las suposiciones introducidas, a saber: la probabilidad enfermedad y las utilidades, de falso y verdadero positivo y falso y verdadero negativo, que también puede variar.

En este caso, con una prevalencia del 10%, la utilidad del criterio 1 es 0,913, la del criterio 2 es 0,949, la del 3, 0,957 y la del 4, 0,959. Basta que la prevalencia sea alta, el 50%, para que el criterio 2 sea el mejor; sin embargo, si la prevalencia es tan baja como 1 por mil, el mejor criterio es el 4. Es lógico, en altas prevalencias no importa tanto la tasa de falsos positivos pero sí en bajas. Por eso el criterio de positividad de la tuberculina o el descenso del ST es diferente según la probabilidad de enfermedad.

¿Cuál es el umbral de probabilidad de enfermedad que hace que tratar sea mejor que no tratar? Evidentemente el umbral se sitúa donde (4) y (5) sean iguales, cuando (4) sea mejor que (5) conviene tratar, y viceversa. Igualando las dos ecuaciones y despejando pE, se obtiene (4), que es el punto de corte óptimo de la curva ROC.

En el numerador se resta a las utilidades de un verdadero negativo (UVN) lo que se pierde por los falsos positivos (UFP) al tratar sanos; son los efectos perjudiciales del test: el coste (C). Y en denominador es el beneficio de encontrar enfermos (B). Se puede reescribir la ecuación (4). Las ecuaciones (6) y (7) representan el punto de la curva ROC óptimo para esas suposiciones, que coincide con el LR ideal, como se ha visto antes (tabla 3).

Supongamos (modificado de Sox HC et al: Medical Decisión Making. Boston 1988; Butterworths) una prueba serológica que sirve para diagnosticar una infección grave que se emplea en el diagnóstico de un caso cuya probabilidad pre-test es del 0.5 y en la detección de infección en personas asintomáticas en un banco de sangre, siendo la probabilidad pre-test del 0.001. El tratamiento incurre en costes importantes al ser muy agresivo y tener escaso beneficios pues es una enfermedad poco vulnerable a la terapia: la relación coste/beneficio es 2/1. Sin embargo, detectar portadores tiene grandes beneficios y bajos costes: 1/50. ¿Cómo definir para cada caso la positividad?

La ecuación (7) quedaría así: $2/1 \cdot 0.50/0.5 = 2$. Tendremos que encontrar un valor de la prueba que cumpla que la sensibilidad sea el doble de la tasa de falsos positivos, un LR de 2. En el caso de las pruebas de cribado la ecuación (7) quedaría así: $1/50 \cdot 0.999/0.001 = 20$. Se debe hallar un valor del test que tenga un LR de 20 para que sea útil como prueba de screening. Es decir, la especificidad debe primar aquí.

En resumen, conviene tener flexibilidad a la hora de decidir el criterio de anormalidad de una prueba. No es correcto juzgar las alteraciones electrocardiográficas sin una referencia poblacional: ¿en qué grupo de riesgo coloco a este paciente? (probabilidad a priori) ¿qué beneficios espero obtener del tratamiento si lo califico como enfermo? ¿En qué costes incurro si a un sano lo trato como enfermo? ¿Y si no trato a un enfermo?. La sensibilidad y especificidad son características de la prueba, pero el acto médico es específico del paciente.

Efectivamente, el viejo aforismo no existen enfermedades sino enfermos, que puede servir para justificar cualquier acción ante el evaluador de la práctica clínica, cobra relevancia tras estas reflexiones. Aunque pensamos en enfermedades, como una constelación de signos, síntomas y pruebas complementarias que con mayor probabilidad tiene una determinada entidad fisiopatológica en contraste con otras, lo que diagnosticamos y tratamos es un enfermo. Conviene, por tanto, personalizar lo más posible las herramientas que empleamos. Sé que puede resultar antipático hacer ejercicios como los que he propuesto. Pero creo que resulta interesante examinar las suposiciones, de manera que podamos analizar su robustez, haciéndolas variar dentro de los límites sensatos. Uno se puede encontrar con sorpresas.

Tabla 4: Utilidades, en el supuesto de la tabla 1, para distintas prevalencias

pE	1-pE	NIVEL	S	1-S	E	1-E	UVP	UVN	UFP	UFN	UE	UNE	U
0,1	0,9	1	0,99	0,01	0,1	0,9	0,7	1	0,93	0,5	0,069	0,843	0,913
		2	0,90	0,1	0,70	0,3					0,068	0,881	0,949
		3	0,70	0,3	0,90	0,1					0,064	0,893	0,957
		4	0,50	0,5	0,99	0,01					0,06	0,899	0,959
0,5	0,5	1	0,99	0,01	0,1	0,9					0,349	0,468	0,817
		2	0,90	0,1	0,70	0,3					0,34	0,489	0,829
		3	0,70	0,3	0,90	0,1					0,32	0,496	0,816
		4	0,50	0,5	0,99	0,01					0,3	0,499	0,799
0,001	0,999	1	0,99	0,01	0,1	0,9					0,00069	0,936	0,936
		2	0,90	0,1	0,70	0,3					0,00068	0,978	0,978
		3	0,70	0,3	0,90	0,1					0,00064	0,992	0,992
		4	0,50	0,5	0,99	0,01					0,006	0,998	0,998

Palabras clave:

De la evidencia científica a las decisiones clínicas

e-notas de evaluación

Autores/as: Martín Caicoya Gómez-Morán

Título: Cuando pedir una prueba diagnóstica

Nº 3 de 2005

Disponible en: [Enlace a la noticia](#)