



FORMACIÓN MÉDICA

El dilema del vigilante. Valoración de pruebas diagnósticas

Molina Arias M.

Hospital Infantil Universitario La Paz, Madrid.

Resumen

Existen una serie de parámetros que caracterizan las pruebas diagnósticas, como son la sensibilidad, la especificidad, los valores predictivos y los cocientes de probabilidad. Solo estos últimos nos permiten el cálculo de la probabilidad del paciente de presentar la enfermedad, con independencia de la prevalencia en la población. Para caracterizar las pruebas con resultado cuantitativo se utilizan las curvas ROC, cuyo parámetro más significativo es el área bajo la curva.

Introducción

Existen una serie de parámetros que caracterizan las pruebas diagnósticas, como son la sensibilidad, la especificidad, los valores predictivos y los cocientes de probabilidad. Solo estos últimos nos permiten el cálculo de la probabilidad del paciente de presentar la enfermedad, con independencia de la prevalencia en la población. Para caracterizar las pruebas con resultado cuantitativo se utilizan las curvas ROC, cuyo parámetro más significativo es el área bajo la curva.

El mundo de la medicina es un mundo de incertidumbre. Nunca podemos estar seguros de nada al 100%, por muy evidente que parezca un diagnóstico, pero no podemos dar palos a diestro y siniestro con técnicas diagnósticas o tratamientos ultramodernos (y nunca inocuos) a la hora de tomar las decisiones que continuamente nos persiguen en nuestra práctica diaria. Es por esto que siempre estamos inmersos en un mundo de probabilidades, donde las certezas son

casi tan infrecuentes como el **mal llamado sentido común** que, como casi todo el mundo sabe, es el menos común de los sentidos.

Imaginemos que estamos en la consulta y acude un paciente que viene porque le han dado una patada en el culo, bastante fuerte, eso sí. Como buenos médicos que somos le preguntamos aquello de ¿qué le pasa?, ¿desde cuándo? y ¿a qué lo atribuye? Y procedemos a una exploración física completa, descubriendo con horror que tiene un hematoma en la nalga derecha. Aquí, amigos míos, las posibilidades diagnósticas son numerosas, así que lo primero que vamos a hacer es un **diagnóstico diferencial exhaustivo**.

Para ello, podremos adoptar cuatro enfoques diferentes. El primero es el enfoque posibilista, que enumerará todos los posibles diagnósticos y tratará de descartar todos ellos de forma simultánea solicitando las pruebas diagnósticas pertinentes. El segundo es el enfoque probabilístico, que ordenará los diagnósticos según su probabilidad relativa y actuará en consecuencia.

Parece un hematoma postraumático (el conocido como síndrome de la patada en el culo), pero alguien podría pensar que la patada no ha sido tan fuerte, así que igual el pobre paciente tiene algún trastorno de coagulación o una discrasia sanguínea con una trombopenia secundaria o, incluso, una enfermedad inflamatoria intestinal con manifestaciones extraintestinales atípicas y fragilidad vascular glútea.

También podríamos utilizar un enfoque pronóstico y tratar de demostrar o descartar la existencia de los diagnósticos posibles con peor pronóstico, con lo que el diagnóstico de síndrome de la patada en el culo perdería interés y nos iríamos a descartar una leucemia crónica. Por último, podría utilizarse un enfoque pragmático, prestando especial interés en descartar primero aquellos diagnósticos que tienen un tratamiento más eficaz (volveríamos a la patada).

Parece que lo más correcto es utilizar una combinación juiciosa de los enfoques probabilístico, pronóstico y pragmático. En nuestro caso indagaríamos si la intensidad del traumatismo justifica la magnitud del hematoma y, en ese caso, indicaríamos unos paños calientes y nos absteríamos de realizar más pruebas diagnósticas. Y este ejemplo parece un delirio mío, pero os puedo asegurar que conozco gente que hace la lista completa y tira de prueba diagnóstica ante cualquier sintomatología, sin reparar en gastos ni riesgos. Y, además, alguno que yo me sé pensaría en alguna otra posibilidad más exótica que no acabo de imaginar y aún el paciente tendría que estar agradecido si su diagnóstico no precisa de la realización de una esfinterotomía anal forzada. Y es que, como ya hemos comentado, la lista de espera para obtener un poco de sentido común supera en muchas ocasiones a la lista de espera quirúrgica.

Imaginad ahora otro paciente con un complejo sintomático menos estúpido y absurdo que el del ejemplo previo. Por ejemplo, un niño con síntomas de enfermedad celiaca. Antes de que realicemos ninguna prueba diagnóstica, nuestro paciente ya tiene una probabilidad de padecer la enfermedad. Esta probabilidad vendrá condicionada por la prevalencia de la enfermedad en la población de la que procede y es lo que se denomina probabilidad preprueba. Esta probabilidad se encontrará en algún punto en relación con dos umbrales que os muestro en la figura 1: el umbral de diagnóstico y el umbral terapéutico.

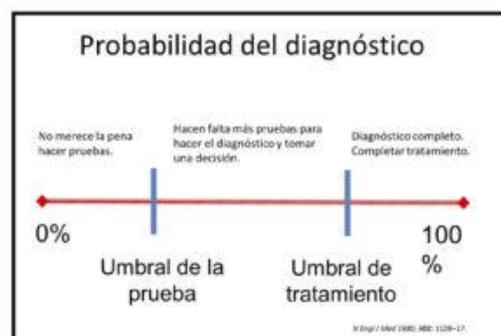


Figura 1. Umbrales

Lo habitual es que la probabilidad preprueba de nuestro paciente no nos permita ni descartar la enfermedad con una seguridad razonable (tendría que ser muy baja, por debajo del umbral diagnóstico) ni confirmarla con la seguridad suficiente como para iniciar el tratamiento (tendría que estar por encima del umbral terapéutico).

Realizaremos entonces la prueba que consideremos indicada, obteniendo una nueva probabilidad de enfermedad según el resultado que nos dé, la llamada probabilidad postprueba. Si esta probabilidad es tan alta como para realizar el diagnóstico e iniciar el tratamiento habremos cruzado el umbral terapéutico. Ya no hará falta realizar pruebas adicionales, ya que tendremos la certeza suficiente para asegurar el diagnóstico y tratar al paciente, siempre

dentro de los rangos de incertidumbre de nuestro oficio.

¿Y de qué depende nuestro umbral de tratamiento? Pues hay varios factores implicados. Cuánto mayor riesgo, coste o efectos adversos tenga el tratamiento en cuestión, mayor será el umbral que exigiremos para tratar. Por otra parte, cuanto mayor gravedad comporte omitir el diagnóstico, menor será el umbral terapéutico que aceptaremos.

Pero puede ocurrir que la probabilidad postprueba sea tan baja que nos permita descartar la enfermedad con una seguridad razonable. Habremos cruzado entonces el umbral de diagnóstico, también llamado umbral negativo de prueba. Es evidente que, en esta situación, no estará indicado realizar más pruebas diagnósticas y, mucho menos, iniciar el tratamiento. Sin embargo, en muchas ocasiones el cambio de probabilidad de preprueba a postprueba nos sigue dejando en tierra de nadie, sin alcanzar ninguno de los dos umbrales, por lo que nos veremos obligados a realizar pruebas adicionales hasta que alcancemos uno de los dos límites.

Y esta es nuestra necesidad de todos los días: conocer la probabilidad postprueba de nuestros pacientes para saber si descartamos o confirmamos el diagnóstico, si dejamos al paciente tranquilo o le fustigamos con nuestros tratamientos. Y es que el planteamiento simplista de que un paciente está enfermo si la prueba diagnóstica es positiva y sano si es negativa es totalmente erróneo, por más que sea la creencia generalizada entre aquellos que indican las pruebas. Tendremos que buscar, pues, algún parámetro que nos indique qué utilidad puede tener una prueba diagnóstica determinada para servir para el fin que necesitamos: saber la probabilidad de que el paciente tenga la enfermedad.

Y esto me recuerda el enorme problema que me consultó el otro día un cuñado. El pobre hombre está muy preocupado con un dilema que le ha surgido. Resulta que va a montar un pequeño comercio y quiere contratar un vigilante para ponerlo en la puerta y que detecte a los que se llevan algo sin pagar. Y el problema es que tiene dos candidatos y no sabe por cuál decidirse. Uno de ellos para a casi todo el mundo, con lo que no se le escapa ningún chorizo. Eso sí, mucha gente honrada se ofende cuando se le pide que abra el bolso antes de salir y lo mismo la próxima vez se va a comprar a otro sitio. El otro es todo lo contrario: no para a casi nadie pero, eso sí, si para a uno, seguro que lleva algo robado. Este ofende a pocos honrados, pero se le escapan demasiados chorizos. Dificil decisión...

¿Y por qué me viene a mí mi cuñado con este cuento? Pues porque sabe que yo me enfrento a diario con un dilema similar cada vez que tengo que elegir una prueba diagnóstica para saber si un paciente está enfermo y le tengo que tratar. Ya hemos dicho que **el positivo de una prueba no nos asegura el diagnóstico**, al igual que la pinta de chorizo no asegura que el pobre cliente nos haya robado.

Veámoslo con un ejemplo. Cuando queremos saber el valor de una prueba diagnóstica, habitualmente comparamos sus resultados con los de un patrón de referencia o patrón oro (el gold standard de los que saben inglés), que es una prueba que, idealmente, es siempre positiva en los enfermos y negativa en los sanos. Ahora supongamos que yo hago un estudio en mi consulta del hospital con una prueba diagnóstica nueva para detectar una determinada enfermedad y obtengo los resultados de la tabla adjunta (los enfermos son los que tienen la prueba de referencia positiva y los sanos, negativa).

Empecemos por lo fácil. Tenemos 1598 sujetos, 520 de ellos enfermos y 1078 sanos. La prueba nos da 446 positivos, 428 verdaderos (VP) y 18 falsos (FP). Además, nos da 1152 negativos, 1060 verdaderos (VN) y 92 falsos (FN). Lo primero que podemos determinar es la capacidad de la prueba para distinguir entre sanos y enfermos, lo que me da pie para introducir los dos primeros conceptos: sensibilidad (S) y especificidad (E). La S es la probabilidad de que la prueba clasifique correctamente a los enfermos o, dicho de otro modo, la probabilidad de que el enfermo sea positivo. Se calcula dividiendo los VP por el número de enfermos. En nuestro caso es de 0,82 (voy a emplear tantos por uno, pero si a alguien le gustan más los porcentajes ya sabe: a multiplicar por 100). Por otra parte, la E es la probabilidad de que se clasifique correctamente a los sanos o, dicho de otro modo, de que los sanos tengan un resultado negativo. Se calcula dividiendo los VN entre el número de sanos. En nuestro ejemplo, 0,98. Alguien podrá pensar que ya tenemos medido el valor de la nueva prueba, pero no hemos hecho nada más que empezar. Y esto es así porque S y E nos miden de alguna manera la capacidad de la prueba para discriminar sanos de enfermos, pero nosotros lo que en realidad necesitamos saber es la probabilidad de que un positivo sea enfermo y de que un negativo sea sano y, aunque puedan parecer conceptos similares, en realidad son bien diferentes.

La posibilidad de que un positivo sea enfermo se conoce como valor predictivo positivo (VPP) y se calcula dividiendo el número de enfermos con prueba positiva entre el número total de positivos. En nuestro caso es de 0,96. Esto sí quiere decir que un positivo tiene un 96% de probabilidad de estar enfermo. Por otra parte, la probabilidad de que un negativo sea sano se expresa

mediante el valor predictivo negativo (VPN), que es el cociente de sanos con resultado negativo entre el número total de negativos. En nuestro ejemplo vale 0,92 (un negativo tiene una probabilidad del 92% de estar sano). Esto ya se va pareciendo más a lo que dijimos al principio que necesitábamos: la probabilidad postprueba de que el paciente esté realmente enfermo. Y ahora es cuando las neuronas empiezan a recalentarse. Resulta que S y E son dos características intrínsecas de la prueba diagnóstica. Los resultados serán los mismos siempre que hagamos la prueba en unas condiciones similares, con independencia de a quién se la hagamos. Pero esto no es así con los valores predictivos, que varían según la prevalencia de la enfermedad en la población en la que hacemos la prueba. Esto quiere decir que la probabilidad de que un positivo esté enfermo depende de lo frecuente o rara que sea la enfermedad en su población. Sí, sí, habéis leído bien: la misma prueba positiva expresa diferente riesgo de estar enfermo, y, para los incrédulos, os pongo otro ejemplo. Supongamos que esta misma prueba la hace un coleguilla mío en su consulta del Centro de Salud, donde la población es proporcionalmente más sana (esto es lógico, todavía no han pasado por el hospital). Si veis los resultados de la tabla, y os molestáis en calcular, veréis que obtiene una S de 0,82 y una E de 0,98, lo mismo que me salía a mí en mi consulta. Sin embargo, si calculáis los valores predictivos, veréis que el VPP es de 0,9 y el VPN de 0,95. Y esto es así porque las prevalencias de la enfermedad (enfermos/totales) son distintas en las dos poblaciones: 0,32 en mi consulta de hospital y 0,19 en la suya. O sea, que en los casos de prevalencia más alta un positivo ayuda más para confirmar la enfermedad y un negativo ayuda menos para descartarla. Y al revés, si la enfermedad es muy rara un negativo permitirá descartar la

enfermedad con una seguridad razonable, pero un positivo nos ayudará mucho menos a la hora de confirmarla.

HOSPITAL			
	Enfermos	Sanos	
Positivos	428	18	446
Negativos	52	1060	1152
	520	1078	1598

$S = \text{Enf Pos/Enf} = 428/520 = 0,82$
 $E = \text{San Neg/San} = 1060/1078 = 0,98$
 $VPP = \text{Enf Pos/Pos} = 428/446 = 0,96$
 $VFN = \text{San Neg/Neg} = 1060/1152 = 0,92$
 $RVP = S/(1-E) = 41$
 $RVN = (1-S)/E = 0,18$
 $\text{Prevalencia} = \text{Enf/Total} = 0,32$
 $OPos = \text{Prev}/(1-Prev) = 0,47$
 $OPos = OPos/RVP = 19,27$
 $\text{ProbPos} = OPos/(OPos+1) = 0,95$
 $CP = \text{ProbPos/Prev} = 2,98$

CENTRO DE SALUD			
	Enfermos	Sanos	
Positivos	200	20	220
Negativos	44	1000	1044
	244	1020	1264

$S = \text{Enf Pos/Enf} = 200/244 = 0,82$
 $E = \text{San Neg/San} = 1000/1020 = 0,98$
 $VPP = \text{Enf Pos/Pos} = 200/220 = 0,9$
 $VFN = \text{San Neg/Neg} = 1000/1044 = 0,95$
 $RVP = S/(1-E) = 41$
 $RVN = (1-S)/E = 0,18$
 $\text{Prevalencia} = \text{Enf/Total} = 0,19$
 $OPos = \text{Prev}/(1-Prev) = 0,23$
 $OPos = OPos/RVP = 9,43$
 $\text{ProbPos} = OPos/(OPos+1) = 0,9$
 $CP = \text{ProbPos/Prev} = 4,7$

Tabla 1. Valor predictivo y prevalencia

Vemos pues que, como pasa casi siempre en medicina, nos movemos en el poco firme terreno de las probabilidades, ya que **todas** (absolutamente todas) **las pruebas diagnósticas son imperfectas y cometen errores a la hora de clasificar sanos y enfermos**. Entonces, ¿cuándo merece la pena utilizar una prueba determinada? Pues si pensamos que un determinado sujeto tiene ya una probabilidad de estar enfermo antes de hacerle la prueba (la prevalencia de la enfermedad en su población), solo nos interesará utilizar pruebas que aumenten esa probabilidad lo suficiente como para justificar el inicio del tratamiento pertinente (en otro caso tendríamos que hacer otra prueba hasta alcanzar el nivel umbral de probabilidad que justifique el tratamiento).

Y aquí es donde el tema se empieza a poner antipático. El cociente de probabilidad positivo (CPP) o razón de verosimilitud positiva nos indica cuánto más probable es tener un positivo en un enfermo que en un sano. La proporción de positivos en los enfermos es la S. La proporción de los positivos en sanos son los FP, que serían aquellos sanos que no dan negativo o, lo que es lo mismo, 1-E. Así, el $CPP = S / (1-E)$. En nuestro caso (del hospital) vale 41 (el mismo aunque utilicemos porcentajes para S y E). Esto puede interpretarse como que es 41 veces más probable encontrar un

resultado positivo en un enfermo que en un sano.

Puede calcularse también el CPN (el negativo), que expresa cuánto más probable es encontrar un negativo en un enfermo que en un sano. Los enfermos negativos son aquellos que no dan positivo (1-S) y los sanos negativos son los VN (la E de la prueba). Luego el $CPN = (1-S)/E$. En nuestro ejemplo 0,18.

Un cociente de probabilidad igual a 1 indica que el resultado de la prueba no modifica la probabilidad de estar enfermo. Si es mayor que 1 aumenta esta probabilidad y, si es menor, la disminuye. Este parámetro es el que usamos para determinar la potencia diagnóstica de la prueba. Valores >10 para CPP (o <0,1 para CPN) indican que se trata de una prueba muy potente que apoya (o contradice) fuertemente el diagnóstico; de 5-10 (o de 0,1-0,2) indican poca potencia de la prueba para apoyar (o descartar) el diagnóstico; de 2-5 (o de 0,2-0,5) indican que la aportación de la prueba es dudosa; y, por último, de 1-2 (o de 0,5-1) indican que la prueba no tiene utilidad diagnóstica.

El cociente de probabilidad no expresa una probabilidad directa, pero nos sirve para calcular las probabilidades de ser enfermo antes y después de dar positivo en la prueba diagnóstica por medio de la regla de Bayes, que dice que la odds postprueba es igual al producto de la odds preprueba por el cociente de probabilidad. Para transformar la prevalencia en odds preprueba usamos la fórmula $\text{odds} = p/(1-p)$. En nuestro caso valdría 0,47. Ahora ya podemos calcular la odds postprueba (OPos) multiplicando la preprueba por el cociente de probabilidad. En nuestro caso, la odds postprueba positiva vale 19,27. Y por último, transformamos la odds postprueba en probabilidad

postprueba usando la fórmula $p = \text{odds}/(\text{odds}+1)$. En nuestro ejemplo vale 0,95, lo que quiere decir que si nuestra prueba es positiva la probabilidad de estar enfermo pasa de 0,32 (la prevalencia o probabilidad preprueba) a 0,95 (probabilidad posprueba).

Si todavía queda alguien leyendo a estas alturas, le diré que no hace falta saberse todo este galimatías de fórmulas. Existen en Internet múltiples páginas con calculadoras para obtener todos estos parámetros a partir de la tabla 2×2 inicial con un esfuerzo miserable. Además, la probabilidad postprueba puede calcularse de forma sencilla utilizando el nomograma de Fagan (figura 2). Este gráfico representa en tres líneas verticales de izquierda a derecha la probabilidad preprueba (se representa invertida), el cociente de probabilidades y la probabilidad postprueba resultante.

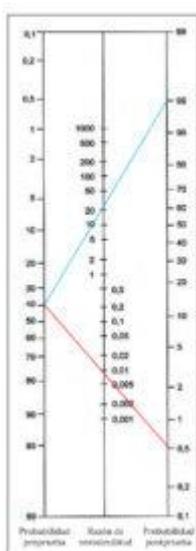


Figura 2. Nomograma Fagan

Para calcular la probabilidad postprueba tras un resultado positivo, trazamos una línea desde la prevalencia (probabilidad preprueba) hasta el CPP y la prolongamos hasta el eje de la probabilidad postprueba. De modo similar, para calcular la probabilidad

postprueba tras un resultado negativo, prolongaríamos la línea que une la prevalencia con el valor del CPN. De esta manera, con esta herramienta podemos calcular de modo directo la probabilidad postprueba conociendo los cocientes de probabilidades y la prevalencia. Además, podremos utilizarlo en poblaciones con distintas prevalencias, simplemente modificando el origen de la línea en el eje de la probabilidad preprueba.

Hasta aquí ya hemos definido los parámetros que nos sirven para cuantificar la potencia de una prueba diagnóstica y hemos visto las limitaciones de sensibilidad, especificidad y valores predictivos y como los más útiles de forma general son los cocientes de probabilidades. Pero, os preguntaréis, ¿qué es bueno?, ¿qué sea sensible?, ¿Qué sea específica?, ¿las dos cosas? Aquí vamos a volver al dilema del vigilante que se le ha planteado a mi pobre cuñado, que le hemos dejado abandonado, porque todavía no hemos respondido cuál de los dos vigilantes le aconsejamos que contrate, el que para a casi todo el mundo para mirarle el bolso y ofende a mucha gente que no roba nada, o el que no para a casi nadie pero tampoco falla con el que para, aunque se escapen muchos ladrones. ¿Y cuál creéis que es mejor de los dos? La respuesta es muy sencilla: depende.

Los que todavía estéis despiertos a estas alturas ya os habréis dado cuenta de que el primer vigilante (el que registra a muchos) es, sin ánimo de ofender, el sensible, mientras que el segundo es el específico. ¿Qué nos interesa más, que el vigilante sea sensible o específico? Pues depende, por ejemplo, de donde tengamos el comercio. Si lo hemos abierto en un barrio de gente bien, no nos interesará mucho el primero, ya que, en realidad, poca gente robará y nos interesa más no ofender a los

clientes para que no se vayan. Pero si ponemos la tienda en frente de la Cueva de Alí-Babá sí que nos traerá más cuenta contratarle para que nos detecte el mayor número posible de clientes que se llevan género robado. Pero también puede depender de lo que vendamos en la tienda. Si tenemos un “todo a un euro” (o un “todo a cien” para los nostálgicos) podemos contratar al vigilante específico, aunque se nos escape alguno (total, perderemos poco dinero). Pero si vendemos joyería fina no queremos que se escape ningún ladrón y contrataremos al sensible (preferiremos que alguien inocente se moleste por ser registrado a que se nos escape uno con un diamante de los gordos).

Pues esto mismo ocurre en medicina con la elección de las pruebas diagnósticas: tendremos que decidir en cada caso si nos interesa más una sensible o una específica, porque no siempre las pruebas disponibles tienen un alto valor de estos dos parámetros. En general, se prefiere una prueba sensible cuando los inconvenientes de obtener falsos positivos (FP) son menores que los de los falsos negativos (FN). Por ejemplo, supongamos que vamos a vacunar a un grupo de enfermos y sabemos que la vacuna es letal en los que tienen determinado error metabólico. Es claro que nos interesará que no se escape ningún enfermo sin diagnosticar (que no haya FN), aunque no pasa nada si a algún sano le etiquetamos de tener el error metabólico (un FP): será preferible no vacunar a un sano por pensar que tiene la metabolopatía (aunque no la tenga) que cargarnos a uno con la vacuna por pensar que no la tenía. Otro ejemplo menos dramático: en medio de una epidemia nos interesará una prueba muy sensible para poder aislar al mayor número posible de enfermos. El problema aquí es el de los desgraciados sanos positivos (FP) que meteríamos

con los infectados, a los cuáles haríamos un flaco favor con la maniobra. Claro que bien podríamos hacer, a todos los positivos de la primera prueba, una segunda de confirmación que sea muy específica para evitar este calvario a los FP. Por otra parte, se prefiere una prueba específica cuando es mejor tener FN que FP, como cuando queremos estar seguros de que un enfermo realmente lo está. Imaginemos que el resultado positivo de una prueba conlleva un tratamiento consistente en una operación quirúrgica: nos convendrá bastante estar seguros de que no vamos a operar a ningún sano. Otro ejemplo es el de las enfermedades cuyo diagnóstico puede ser muy traumático para el paciente y que encima son prácticamente incurables o no tienen tratamiento. Aquí primaremos la especificidad para no darle un disgusto innecesario a ningún sano. Por el contrario, si la enfermedad es muy grave pero tiene tratamiento, probablemente prefiramos una prueba sensible.

Hasta aquí hemos hablado de pruebas con resultado dicotómico: positivo o negativo. Pero, ¿qué pasa cuando el resultado es cuantitativo? Imaginemos que medimos la glucemia en ayunas. Debemos decidir hasta qué valor de glucemia consideramos normal y por encima de cuál nos parecerá patológico. Y esta es una decisión crucial, porque S y E dependerán del punto de corte que elijamos.

Para ayudarnos a elegir disponemos de la curva de características operativas para el receptor, mundialmente conocida como curva ROC (*receiver operating characteristic*).

Representamos en ordenadas (eje y) la S y en abscisas el complementario de la E (1-E) y trazamos una curva en la que cada punto de corte representa la

probabilidad de que la prueba clasifique correctamente a una pareja sano-enfermo tomada al azar. La diagonal del gráfico representaría la “curva” si la prueba no tuviese capacidad ninguna de discriminar sanos de enfermos. Como veis en la figura 3, la curva suele tener un segmento de gran pendiente donde aumenta rápidamente la S sin que apenas varíe la E: si nos desplazamos hacia arriba podemos aumentar la S sin que prácticamente nos aumenten los FP.

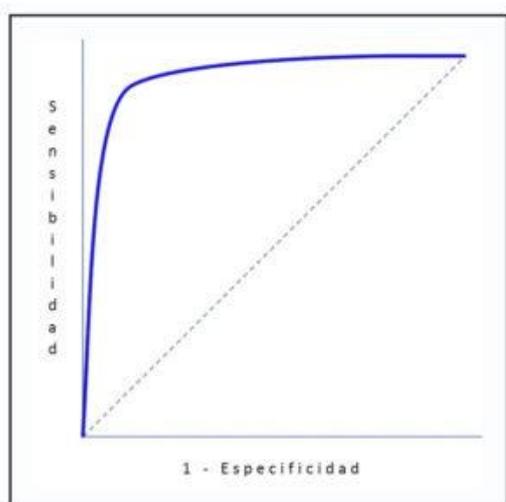


Figura 3. ROC

Pero llega un momento en que llegamos a la parte plana. Si seguimos desplazándonos hacia la derecha llegará un punto a partir del cual la S ya no aumentará más, pero comenzarán a aumentar los FP. Si nos interesa una prueba sensible, nos quedaremos en la primera parte de la curva. Si queremos especificidad tendremos que irnos más hacia la derecha. Y, por último, si no tenemos predilección por ninguna de las dos (nos preocupa igual obtener FP que FN), el mejor punto de corte será el más próximo al ángulo superior izquierdo. Para esto, algunos utilizan el denominado índice de Youden, que es el que optimiza al máximo los dos parámetros y que se calcula sumando S y E y restando 1. Cuanto más alto, menos pacientes mal clasificados por la prueba diagnóstica.

Un parámetro de interés es el área bajo la curva (ABC), que nos representa la probabilidad de que la prueba diagnóstica clasifique correctamente al paciente al que se le practique (figura 4).

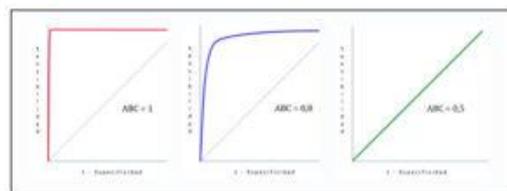


Figura 4. Curvas ROC

Una prueba ideal con S y E del 100% tiene un área bajo la curva de 1: siempre acierta. En clínica, una prueba cuya curva ROC tenga un $ABC > 0,9$ se considera muy exacta, entre 0,7-0,9 de exactitud moderada y entre 0,5-0,7 de exactitud baja. En la diagonal el ABC es igual a 0,5 e indica que da igual hacer la prueba que tirar una moneda al aire para decidir si el paciente está enfermo o no. Valores por debajo de 0,5 indican que la prueba es incluso peor que el azar, ya que clasificará sistemáticamente a enfermos como sanos y viceversa. Curiosas las curvas ROC, ¿verdad? Pues su utilidad no se limita a la valoración de la bondad de las pruebas diagnósticas con resultado cuantitativo. Las curvas ROC sirven también para determinar la bondad del ajuste de un modelo de regresión logística para predecir resultados dicotómicos, pero esa es otra historia...

Bibliografía

1. Molina Arias M, Ochoa Sangrador C. Evaluación de la validez de las pruebas diagnósticas (I). Sensibilidad. Especificidad. Evid Pediatr. 2016;12:34. ([PDF](#))
2. Molina Arias M, Ochoa Sangrador C. Evaluación de la validez de las pruebas diagnósticas (II). Valores predictivos. Evid Pediatr. 2016;12:53. ([PDF](#))
3. Molina Arias M, Ochoa Sangrador C. Evaluación de la validez de las pruebas

diagnósticas (III). Valores predictivos. Evid Pediatr. 2016;12:69. ([PDF](#))

4. Molina Arias M, Ochoa Sangrador C. Pruebas diagnósticas con resultados continuos o politómicos. Curvas ROC. Evid Pediatr. 2017;13:12. ([PDF](#))

Correspondencia al autor

Manuel Molina Arias

mma1961@gmail.com

Servicio de Gastroenterología

Hospital Infantil Universitario La Paz, Madrid.

Aceptado para blog en marzo de 2018.