

Hay que saber lo que se pide

A diario encontramos artículos que nos muestran nuevas pruebas diagnósticas que parecen haber sido diseñadas para solucionar todos nuestros problemas. Pero no debemos caer en la tentación de hacer caso a todo lo que leamos sin recapacitar antes un poco en lo que hemos leído. Al fin y al cabo, si hiciésemos caso a todo lo que leemos estaríamos hinchados de beber Coca-Cola.



Ya sabemos que una prueba diagnóstica no nos va a decir si una persona está o no enferma. Su resultado únicamente nos permitirá aumentar o disminuir la probabilidad de que el individuo esté enfermo o no, de forma que nosotros nos atreveremos a confirmar o descartar el diagnóstico, pero siempre con cierto grado de **incertidumbre**. Cualquiera tiene cierto riesgo de

padecer cualquier enfermedad, que no es más que la **prevalencia** de la enfermedad en la población general. Pero si, además de pertenecer a la población, uno tiene la desgracia de tener síntomas, esa probabilidad irá aumentando hasta alcanzar un primer umbral en el que se justifique realizar pruebas diagnósticas. La utilidad de la prueba diagnóstica estará en su capacidad para disminuir la probabilidad por debajo de este umbral (y descartar el diagnóstico) o, por el contrario, en aumentarla hasta el umbral en el que se justifique iniciar el tratamiento. Claro que a veces la prueba nos deja a medio camino y tenemos que hacer pruebas adicionales antes de confirmar el diagnóstico con la seguridad suficiente como para comenzar el tratamiento.

Los estudios de pruebas diagnósticas deben proporcionarnos información sobre la capacidad de una prueba para producir los mismos resultados cuando se realiza en condiciones similares (**fiabilidad**) y sobre la exactitud con la que las mediciones reflejan aquello que miden (**validez**). Pero, además, deben darnos datos sobre su capacidad discriminativa (**sensibilidad y especificidad**), su rendimiento clínico (**valor predictivo positivo y valor predictivo negativo**) y sobre otros aspectos que nos permitan valorar si nos va a merecer la pena practicarla en nuestros pacientes. Y para comprobar si un estudio nos proporciona la información adecuada tenemos que hacer una lectura crítica basada en nuestros **tres pilares**: **validez**, **importancia** y **aplicabilidad**.

Comencemos por la **VALIDEZ**. Lo primero será hacernos unas preguntas básicas de eliminación o criterios primarios sobre el estudio. Si la

respuesta a estas preguntas es no, probablemente lo mejor que podamos hacer es usar el artículo para envolver el bocadillo de media mañana.

¿Se ha comparado la prueba diagnóstica de forma ciega e independiente con un patrón de referencia adecuado?. Hay que revisar que el resultado de la prueba de referencia no se interprete de forma diferente según el resultado de la prueba de estudio, ya que caeríamos en un [sesgo de incorporación](#), que podría invalidar los resultados. Otro problema que puede surgir es que el patrón de referencia tenga muchos resultados poco concluyentes. Si cometemos el error de excluir estos casos dudosos incurriremos en un [sesgo de exclusión de indeterminados](#) que, además de sobrestimar la sensibilidad y la especificidad de la prueba, comprometería la validez externa del estudio, que solo sería aplicable a los pacientes con resultado no dudoso.

¿Los pacientes abarcan un espectro similar al que nos vamos a encontrar en nuestra práctica?. Deben estar claros los criterios de inclusión del estudio, en el que deben participar sanos y enfermos con distinta gravedad o evolución de la enfermedad. Como [ya sabemos](#), la prevalencia influye en el rendimiento clínico de la prueba, con lo que si la validamos, por ejemplo, en un centro terciario (estadísticamente la probabilidad de estar enfermo será mayor) puede sobrestimarse su capacidad diagnóstica si va a utilizarse en un centro de Atención Primaria o en población general (en el que la proporción de enfermos será menor).

Llegados a este punto, si creemos que merece la pena seguir leyendo, pasaremos a los [criterios secundarios](#), que son aquellos que aportan un valor añadido al diseño del estudio. Otra pregunta que debemos hacernos es: ¿influyeron los resultados de la prueba de estudio para decidir si se hacía la de referencia?. Hay que comprobar que no se haya producido un [sesgo de secuencia](#) o [sesgo de verificación diagnóstica](#), mediante el cual excluimos a los que tienen la prueba negativa. Aunque esto es habitual en la práctica corriente (empezamos por pruebas sencillas y solo hacemos las caras o las invasoras en los casos positivos), el hacerlo en un estudio de pruebas diagnósticas compromete la validez de los resultados. Ambas pruebas deben hacerse de forma independiente y ciega, de forma que la subjetividad del observador no influya en los resultados ([sesgo de revisión](#) o [sesgo de valoración ciega](#)). Por último, ¿se describe el método con el detalle suficiente para permitir su reproducción?. Debe quedar claro qué se ha considerado normal y anormal y cuáles han sido los criterios para definir la normalidad y la forma de interpretar los resultados de la prueba.

Una vez analizada la validez interna del estudio pasaremos a considerar la [IMPORTANCIA](#) de los datos presentados. El objetivo de un estudio de diagnóstico es determinar la [capacidad de una prueba](#) para clasificar correctamente a los individuos según la presencia o ausencia de enfermedad.

En realidad, y para ser más exactos, queremos saber cómo aumenta la probabilidad de estar enfermo tras el resultado de la prueba ([probabilidad postprueba](#)). Es, por tanto, esencial que el estudio nos informe acerca de la dirección y magnitud de este cambio (preprueba/postprueba), que sabemos depende de las características de la prueba y, en gran medida, de la prevalencia o [probabilidad preprueba](#).

¿Nos presenta el trabajo las razones de verosimilitud o es posible calcularlas a partir de los datos?. Este dato es fundamental, ya que sin él no podemos calcular el [impacto clínico](#) de la prueba de estudio. Hay que tener especial precaución con las pruebas de resultado cuantitativo en las que es el investigador el que establece un [punto de corte](#) de normalidad. Cuando se utilizan [curvas ROC](#) es frecuente desplazar el punto de corte para favorecer la sensibilidad o la especificidad de la prueba, pero tenemos que valorar siempre cómo afecta esta medida a la validez externa del estudio, ya que puede limitar su aplicabilidad a un grupo determinado de pacientes.

¿Son fiables los resultados?. Habrá que determinar si los resultados son reproducibles y cómo pueden verse afectados por variaciones entre diferentes observadores o al repetir la prueba de forma sucesiva. Pero no solo hay que valorar la fiabilidad, sino también cuán precisos son los resultados. El estudio se hace sobre una muestra de pacientes, pero debe proporcionar una estimación de sus valores en la población, por lo que los resultados deben expresarse con sus correspondientes [intervalos de confianza](#).

El tercer pilar de la lectura crítica es el de la [APLICABILIDAD](#) o validez externa, que nos ayudará a determinar si los resultados son útiles para nuestros pacientes. En este sentido, debemos hacernos tres preguntas. ¿Disponemos de esta prueba y es factible realizarla en nuestros pacientes?. Si no disponemos de la prueba lo único que habremos conseguido leyendo el estudio es aumentar nuestros vastos conocimientos. Pero si disponemos de ella debemos preguntarnos si nuestros pacientes cumplirían los criterios de inclusión y exclusión del estudio y, en caso de que no los cumplan, pensar cómo pueden afectar estas diferencias la aplicabilidad de la prueba.

La segunda pregunta es si conocemos la probabilidad preprueba de nuestros pacientes. Si nuestra prevalencia es muy diferente de la del estudio se puede modificar la utilidad real de la prueba. Una solución puede ser hacer un [análisis de sensibilidad](#) valorando cómo se modificarían los resultados del estudio estudiando varios valores de probabilidad pre y postprueba que sean clínicamente razonables.

Por último, deberíamos hacernos la pregunta más importante: ¿la probabilidad postprueba puede hacer cambiar nuestra actitud terapéutica y servir de ayuda para el paciente?. Por ejemplo, si la probabilidad preprueba es muy baja, probablemente la probabilidad postprueba sea también

muy baja y no alcanzará el umbral de justificación terapéutica, con lo que igual no merece la pena gastar dinero y esfuerzos con esa prueba. Por el contrario, si la probabilidad preprueba es muy alta, en algunos casos merecerá la pena tratar sin hacer ninguna prueba, salvo que el tratamiento sea muy costoso o peligroso. Como siempre, en el medio estará la virtud y será en esas zonas intermedias donde más nos podamos beneficiar del uso de la prueba diagnóstica en cuestión. En cualquier caso, no nos olvidemos nunca de nuestro jefe (me refiero al paciente, no al otro): no hay que contentarse solo con estudiar la eficacia o el coste-efectividad, sino que debemos considerar también los riesgos, molestias y preferencias del paciente, así como las consecuencias que le puede acarrear la realización o no de la prueba diagnóstica.

Si me permitís un consejo, cuando estéis valorando un trabajo sobre pruebas diagnósticas os recomiendo el uso de las [plantillas CASPe](#), que podéis descargaros de su página web. Os ayudarán a hacer la lectura crítica de una manera sistemática y sencilla.

Para terminar, comentaros que todo lo dicho hasta ahora vale para los trabajos específicos de pruebas diagnósticas. Sin embargo, la valoración de pruebas diagnósticas puede formar parte de estudios observacionales como los de cohortes o los de casos y controles, que pueden tener alguna peculiaridad en la secuencia de realización y en los criterios de validación de la prueba de estudio y del patrón de referencia, pero esa es otra historia...

No es oro todo lo que reluce

Tengo un cuñado que 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. Difícil 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. Y es que todavía hay quien piensa que si tú pides una prueba y es positiva ya tienes hecho el diagnóstico y, al revés, que, si estás enfermo, para saberlo no hay más que hacerte la prueba. Y las cosas no son, ni muchos menos, tan sencillas, ni es oro todo lo que reluce, ni todo el oro tiene los mismos quilates.

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

HOSPITAL			
	Enfermos	Sanos	
Positivos	428	18	446
Negativos	92	1060	1152
	520	1078	1598

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

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).

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.

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

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

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. La **razón de verosimilitud positiva** (RVP) o cociente de probabilidad positivo 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, la $RVP = S / (1-E)$. 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 la RVN (la **negativa**), 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 la $RVN = (1-S)/E$. 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 (ó $<0,1$) indican que se trata de una prueba muy potente que apoya (o contradice) fuertemente el diagnóstico; de 5-10 (ó de $0,1-0,2$) indican poca potencia de la prueba para apoyar (o descartar) el diagnóstico; de 2-5 (ó de $0,2-0,5$) indican que la aportación de la prueba es dudosa; y, por último, de 1-2 (ó de $0,5-1$) indican que la prueba no tiene utilidad diagnóstica.

La razón de verosimilitud (likelyhood ratio para los ingleses) 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. Podemos calcular la **odds preprueba** (OPre) como la prevalencia dividida por su complementario (cuánto más probable es que esté enfermo a que no lo esté). En nuestro caso valdría 0,47. Por otro lado, la **odds posprueba** (OPos) se calcula como el producto $RVP \times OPre$. En nuestro caso, 19,27. Y por último, siguiendo el mecanismo inverso al usado para obtener la OPre a partir de la prevalencia, la **probabilidad posprueba** (PrPos) sería

igual a $OPos/(OPos+1)$. 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 2x2 inicial con un esfuerzo miserable. Además, la probabilidad postprueba puede calcularse de forma sencilla utilizando el [nomograma de Fagan](#). Lo que tenemos que saber es cómo valorar adecuadamente la información que nos proporciona una prueba diagnóstica para saber si debemos indicarla en razón de su potencia, costes, molestias para el paciente, etc.

Una última cuestión. Llevamos hablando todo el rato de pruebas positivas o negativas, pero cuando el resultado es cuantitativo debemos ser nosotros los que fijemos qué valor consideramos positivo y cuál negativo, con lo que variarán todos los parámetros que hemos visto en esta entrada, sobre todo la S y la E. ¿Y a cuál de las características de la prueba diagnóstica debemos dar prioridad?. Pues eso dependerá de las características de la prueba y del uso que pretendamos darle, pero esa es otra historia...
