

**FORMACIÓN MÉDICA**

Hay que saber lo que se pide. Lectura crítica de estudios sobre pruebas diagnósticas

Molina Arias M.

Hospital Infantil Universitario La Paz, de Madrid.

Resumen

La potencia de una prueba diagnóstica se valora mediante su capacidad para discriminar sanos y enfermos, así como en los cambios de la probabilidad previa de presentar la enfermedad según el resultado de la prueba. La lectura crítica de estos trabajos valorará su validez metodológica, la importancia de los resultados y su aplicabilidad a nuestra práctica clínica. Para ello, habrá que examinar aspectos como las características de la prueba, la prevalencia o el espectro de enfermedad entre los participantes.

Introducción

La potencia de una prueba diagnóstica se valora mediante su capacidad para discriminar sanos y enfermos, así como en los cambios de la probabilidad previa de presentar la enfermedad según el resultado de la prueba. La lectura crítica de estos trabajos valorará su validez metodológica, la importancia de los resultados y su aplicabilidad a nuestra práctica clínica. Para ello, habrá que examinar aspectos como las características de la prueba, la prevalencia o el espectro de enfermedad entre los participantes.

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 leamos estaríamos hinchados de 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. Por debajo de cierto nivel de probabilidad, nos parece tan poco probable que el paciente esté enfermo que le dejamos tranquilos y no le hacemos ninguna prueba diagnóstica (aunque a algunos les cueste mucho refrenar el impulso de pedir siempre algo). Este es el umbral de prueba o diagnóstico.

Pero si, además de pertenecer a la población, uno tiene la desgracia de tener síntomas, esa probabilidad irá aumentando hasta superar este umbral en el que la probabilidad de presentar la enfermedad ya justifica realizar pruebas

diagnósticas. Una vez que tengamos el resultado de la prueba que hayamos elegido, la probabilidad (probabilidad postprueba) habrá cambiado. Puede que haya cambiado a menos y se haya situado por debajo del umbral de prueba, con lo que descartamos el diagnóstico y volvemos a dejar al paciente tranquilo. También puede que supere otro umbral, el terapéutico, a partir del cual la probabilidad de la enfermedad alcanza el nivel suficiente como para no necesitar más pruebas y poder iniciar el tratamiento.

La utilidad de la prueba diagnóstica estará en su capacidad para disminuir la probabilidad por debajo del umbral de prueba (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 discriminatoria (sensibilidad y especificidad), su rendimiento clínico (valor predictivo positivo y valor predictivo negativo), su capacidad de modificar la probabilidad de enfermedad y cambiar nuestra posición entre los dos umbrales (cocientes de probabilidad o verosimilitud), 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 es bien sabido, 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).

Esto puede parecer difícil de comprender, pero si lo pensamos dos veces veréis que no lo es tanto: cuánto más prevalente sea la enfermedad, más probable es que un positivo de la prueba sea verdadero, aunque sea de casualidad. El mérito es tener más verdaderos en poblaciones con prevalencia baja.

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. Como ya hemos comentado, 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 varía 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. Para esto podemos ayudarnos de una herramienta muy útil, en nomograma de Fagan. Este nomograma permite calcular de forma fácil la probabilidad postprueba para diferentes prevalencias (probabilidad preprueba), siempre que conozcamos con cocientes de verosimilitud de la prueba diagnóstica.

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.

Una aclaración para ir acabando: no debemos confundir los estudios de pruebas diagnósticas con los de reglas de predicción diagnóstica. Aunque la valoración es parecida, las reglas de predicción tienen unas características específicas y unos requerimientos metodológicos que deben valorarse de una forma adecuada y que veremos en otra entrada.

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...

Bibliografía

- Ochoa Sangrador C, González de Dios J, Buñuel Álvarez JC. Evaluación de artículos científicos sobre pruebas diagnósticas. Evid Pediatr. 2007; 3:24. ([HTML](#)) ([PDF](#))

- Ibáñez Pradas V, Modesto Alapont V. MBE en Cirugía Pediátrica. Lectura crítica de artículos. Pruebas diagnósticas (I). Cir Pediatr. 2006;19:123-9. ([PDF](#))

- Ibáñez Pradas V, Modesto Alapont V. MBE en Cirugía Pediátrica. Lectura crítica de artículos. Pruebas diagnósticas (II). Cir Pediatr. 2006;19:130-5. ([PDF](#))

Correspondencia al autor

Manuel Molina Arias

mma1961@gmail.com

FEA Anestesiología.

Servicio de Gastroenterología Hospital Infantil Universitario La Paz, de Madrid.

Aceptado para blog en agosto de 2018.