

Introducción a la Medicina Basada en Evidencias

Medwave. Año XI, No. 7, Julio 2011. Open Access, Creative Commons.

Estudios sobre el diagnóstico de las enfermedades

Autor: Miguel Araujo Alonso⁽¹⁾

Filiación:

⁽¹⁾Hospital de Los Andes, Medwave

Correspondencia: miguel.araujo@medwave.cl

doi: 10.5867/medwave.2011.07.5067

Ficha del Artículo

Citación: Araujo M. Estudios sobre el diagnóstico de las enfermedades. *Medwave* 2011 Jul;11(07). doi:

10.5867/medwave.2011.07.5067

Fecha de envío: 22/6/2011

Fecha de aceptación: 24/6/2011

Fecha de publicación: 1/7/2011

Origen: solicitado

Tipo de revisión: sin revisión por pares

Resumen

En la sección Series, Medwave publica artículos relacionados con el desarrollo y discusión de herramientas metodológicas para la investigación clínica, la gestión en salud, la gestión de la calidad y otros temas de interés. En esta edición se presentan dos artículos que forman parte del programa de formación en Medicina Basada en Evidencias que se dicta por e-Campus de Medwave. El artículo siguiente pertenece a la Serie **"Introducción a la Medicina Basada en Evidencias"**.

Introducción

Los estudios de pruebas diagnósticas tienen como objetivos principales: evaluar la capacidad de discriminación de una prueba para establecer el diagnóstico de enfermedad o el estado de avance de la misma, y estimar el efecto del uso del test sobre el manejo clínico y el pronóstico final de los pacientes.

Es posible encontrar diversos diseños de investigación en evaluación de pruebas diagnósticas. El método por excelencia es la comparación de los resultados del test con los de un estándar de referencia en una muestra consecutiva de sujetos con sospecha de enfermedad. Otras opciones incluyen los estudios caso control y el muestreo según los resultados del test.

Objetivos de los estudios de pruebas diagnósticas

Las distintas opciones de diseño existentes para evaluar las pruebas diagnósticas se relacionan, en primer lugar, con los objetivos que se plantea el autor del estudio. Esos objetivos pueden ir desde evaluar las características operacionales del test en condiciones de laboratorio, hasta conocer el impacto que el uso del test tiene sobre los desenlaces clínicos del paciente.

a. En las fases iniciales del desarrollo de la prueba, las preguntas se relacionan con su desempeño "técnico", por ejemplo, en términos de resolución espacial o su capacidad de discriminación *in vitro*.

b. Ya en el terreno clínico, una primera etapa es aquella en que uno se plantea si los resultados de la prueba serán distintos en los pacientes con enfermedad y en las personas sanas. El uso rutinario que hacemos de muchos exámenes, hace que normalmente no nos cuestionemos al respecto. No obstante, más de alguna vez usted se habrá preguntado si determinado hallazgo pudiera ser "normal", es decir, que quizás ocurre con similar frecuencia, tanto en sujetos enfermos o sintomáticos, como en aquellos que no lo están. Sabido es, por ejemplo, que una alta proporción de personas después de cierta edad presenta alteraciones radiológicas de columna que no se manifiestan por ningún síntoma o limitación funcional.

Para contestar este tipo de preguntas se debe comparar el examen en un grupo de pacientes con enfermedad conocida, con otro grupo de sujetos que no la tienen. Estos estudios de pruebas diagnósticas se denominan "caso-control" o "casos y controles".

Si está suficientemente claro que el test tiene un resultado diferente en los sujetos sanos y enfermos, ya sea porque existe un estudio publicado al respecto, o porque resulta obvio que así es, podemos preguntarnos entonces cuál es su desempeño si se aplica en un grupo de pacientes con sospecha de enfermedad; es decir, bajo condiciones similares a las que uno enfrentaría en la práctica.

Este tipo de preguntas se responden con el diseño "clásico" de los estudios de pruebas diagnósticas, contrastando los resultados del test (o prueba índice) con los de un patrón de oro o referencia. Los resultados de estos estudios se expresan con las conocidas medidas de sensibilidad, especificidad, valores predictivos y LR (*likelihood ratio*).

c. Si el test efectivamente permite avanzar en el diagnóstico, la pregunta que sigue es si esa información afecta de modo significativo el manejo del paciente. ¿Cambian las decisiones terapéuticas con ese diagnóstico más preciso? Imagine el caso extremo de un paciente con un accidente vascular cerebral con compromiso neurológico profundo y fuera del alcance quirúrgico. Conocer las características más específicas de la lesión, ¿modifica la conducta terapéutica?

Revise el siguiente resumen: van Lankveld MA, Peeters PH, van Eijkeren MA, Koot VC, Witteveen PO, Mali WP. The value of abdominal CT scans in decision-making during chemotherapy in ovarian cancer. *Med Oncol*. 2004;21(1):41-8.

d. La pregunta final es si la prueba modifica o no el pronóstico del paciente, es decir, si por el hecho de ejecutarla puede esperarse en definitiva un mejor desenlace clínico. En muchos casos esto es obvio y no se requiere hacer un estudio para demostrarlo.

Por la vía de inducir al uso de tratamientos infectivos o peligrosos, un test con excelente capacidad diagnóstica podría resultar en definitiva dañino para los pacientes. Piense por ejemplo en el efecto psicológico que el hallazgo de un sutil "infiltrado pulmonar crónico" (no detectable en la radiografía convencional) podría tener sobre una paciente con disnea psicógena secundaria a un trastorno de ansiedad. O en cuántos pacientes fueron sometidos en exceso a cirugía por una hernia de núcleo pulposo cuando estas lesiones comenzaron a ser fácilmente detectables con las técnicas de imágenes actuales.

Este tipo de preguntas se aborda dando al test la categoría de intervención, y por lo tanto deben ser respondidas idealmente a través de un estudio experimental, es decir, un ensayo clínico aleatorizado, en el que un grupo es asignado a recibir la intervención -en este caso la prueba diagnóstica-, y otro a una intervención alternativa (otra prueba) o ninguna. También se suelen estudiar mediante diseños de cohorte, en los que el uso de la prueba es la variable exposición a evaluar.

Revise el siguiente resumen: Lidman C, Burman LG, Lagergren A, Orqvist A. Limited value of routine microbiological diagnostics in patients hospitalized for community-acquired pneumonia. *Scand J Infect Dis*. 2002;34(12):873-9.

Desde otra perspectiva, debemos tener presente que los objetivos de las pruebas no se restringen simplemente a establecer o descartar la presencia de enfermedad, sino

que abarcan una gama mucho más compleja de aplicaciones clínicas, por ejemplo:

- Pueden constituir un requisito para apoyar la ejecución de un tratamiento (por ejemplo, la coronariografía no sólo indica que existe una estenosis, también permite planificar la cirugía o procedimiento de revascularización).
- Permiten definir el estado de avance de la enfermedad, es decir, aportan información pronóstica, o que sirve de línea basal para el posterior seguimiento del paciente.
- Los test diagnósticos permiten evaluar la respuesta al tratamiento. ¿Ha oído hablar de los "exámenes de control"?

Diseño de un estudio de pruebas diagnósticas con *Gold Standard*

A grandes rasgos, la estructura de un estudio de pruebas diagnósticas clásico es la siguiente:

1. Se selecciona la prueba que se desea evaluar, especificándola lo más posible, de modo que uno sepa bajo qué condiciones se aplicó (tipo de equipo, técnica utilizada, criterios de positividad, etc.) y cualquier interesado pueda reproducirla si lo desea.

2. Se identifica un método diagnóstico, o varios de ellos que en conjunto representen la mejor forma de establecer el diagnóstico real del paciente, al que llamaremos "prueba o patrón de oro (EO) [*gold standard*]" o "estándar de referencia", contra el cual vamos a comparar la prueba que deseamos evaluar. Por ejemplo:

- Los resultados de la cirugía + la anatomía patológica + seguimiento clínico (EO), respecto de un examen de imágenes no invasivo (por ejemplo, la ecografía para el diagnóstico de apendicitis aguda).
- Los resultados de la tomografía axial computarizada (EO), respecto de la radiografía simple de cráneo, para predecir lesión intracraneana en un paciente con trauma encéfalo craneano.
- Una prueba de laboratorio de alta sensibilidad pero más compleja y de alto costo (EO), respecto de un test rápido que puede realizarse mediante tiras reactivas.

3. Se aplica la prueba sobre un grupo amplio de pacientes, que sea lo más representativo posible de todo el espectro de casos (mezcla de casos o *case mix*) sobre los cuales se aplicaría el nuevo test en la práctica. Por ejemplo, si la prueba pretende diagnosticar un cáncer en fases incipientes, habrá que aplicarla en pacientes con distintos niveles de riesgo de presentar ese cáncer y sin evidencias clínicas de él. Si se aplica sólo en pacientes de muy alto riesgo o con cáncer clínicamente evidente, lo más probable es que en el estudio la prueba parezca ser muy sensible y después cuando se aplique en la práctica, la sensibilidad real termine siendo ostensiblemente más baja.

Habitualmente tanto la prueba como el patrón de referencia se aplican transversalmente, es decir dentro de un período de tiempo acotado, sobre una muestra consecutiva de sujetos que se presentan con elementos que hacen sospechar la enfermedad, sin exclusiones arbitrarias. Se asume que una muestra consecutiva es la más representativa de la población en quien se aplicaría la prueba en el contexto real.

Cabe señalar que la terminología para referirse a este tipo de diseño no es homogénea; por ejemplo, cuando la secuencia en la que se realizan las pruebas sobre la serie es tal que los sujetos primero son sometidos a la prueba índice y después al patrón de referencia, algunos autores hablan de "cohorte diagnóstica" (por la idea de seguimiento, desde el primer test hasta el segundo). Y hay otros que, cuando el patrón es complementado con el seguimiento clínico de los pacientes, se refieren al estudio como un corte transversal "extendido".

En general estos estudios se plantean de manera prospectiva, pero también es frecuente observar diseños retrospectivos o bidireccionales. Como siempre, el uso de métodos retrospectivos de recolección de datos agrega un potencial de error, sometiendo al investigador a la calidad y disponibilidad de los datos existentes.

4. Se somete al patrón de oro a todos los pacientes a los que se aplicó la prueba, de manera de estar seguro que los pacientes que resultaron negativos en la prueba no tenían realmente un diagnóstico positivo y al revés, que los que resultaron positivos no eran en realidad falsos positivos.

Esto no siempre se cumple; muchas veces los pacientes que resultan negativos en la prueba no son sometidos al patrón de oro. Suponga que se desea evaluar la utilidad de la Rx de cráneo en pacientes con trauma encéfalo craneano (TEC) leve y su patrón de oro es la tomografía axial computarizada (TAC). En teoría, lo correcto sería someter a todos los pacientes a ambos exámenes, sin embargo pudiera ocurrir que los pacientes con Rx normal y asintomáticos después de haberse realizado la radiografía, no sean finalmente enviados a TAC por el costo que implica esta última y el bajo riesgo de lesión intracraneana que poseen. Si así ocurre, es posible que algunos de esos pacientes sí posean una lesión intracraneana que no fue detectada por el hecho de no haberles realizado el segundo examen (aunque esa lesión no sea clínicamente importante).

Otra condición para asegurar la validez de los resultados es que el patrón de oro sea aplicado por un investigador independiente de quien aplicó la prueba y que además esté ciego a sus resultados, de forma que al interpretar el patrón de oro no se vea influenciado por ella. En nuestro ejemplo, quien interpreta la TAC no debiera conocer los resultados de la Rx.

5. Una vez registrados los resultados de ambas pruebas, la que deseamos evaluar y el patrón de oro (que representa en teoría "el verdadero diagnóstico"), lo que sigue es el análisis de los datos. Para ese efecto, se construye una tabla de 2x2 o "tabla de contingencia", en la forma que se indica a continuación:

Resultado de la prueba	Verdadero diagnóstico	
	Enfermo	Sano
Positivo	Verdaderos Positivos (VP)	Falsos Positivos (FP)
Negativo	Falsos Negativos (FN)	Verdaderos Negativos (VN)

Utilicemos ahora un ejemplo ficticio para ilustrar la forma de calcular los indicadores de resultado básicos de todo estudio de pruebas diagnósticas: sensibilidad, especificidad y valores predictivos. La prueba es la ecografía abdominal y el patrón de oro son: los resultados de la cirugía y la biopsia en los pacientes operados, y el seguimiento en aquéllos que no fueron intervenidos

inicialmente (puede ser que algunos de los pacientes terminen siendo operados gracias al seguimiento, y en ellos el resultado final estará dado igualmente por los resultados de la cirugía y la biopsia; si no son intervenidos y evolucionan favorablemente, se puede asumir que no tenían apendicitis).

Resultado de la ecografía	Resultado de la cirugía / biopsia / seguimiento		
	Apendicitis aguda	Sin apendicitis	Total
Anormal	130	10	140
Normal	20	190	210
Total	150	200	350

Sensibilidad: $\frac{130}{130+20} = 0,87 \text{ (87\%)}$
Especificidad: $\frac{190}{190+10} = 0,95 \text{ (95\%)}$
Valor predictivo positivo: $\frac{130}{130+10} = 0,93 \text{ (93\%)}$
Valor predictivo negativo: $\frac{190}{190+20} = 0,90 \text{ (90\%)}$

Hoy día Internet nos ofrece **calculadoras online** que permiten hacer estos análisis en forma automática.

Estudios sobre tamizaje

El tamizaje implica el uso de una prueba en población asintomática, en la que la prevalencia de enfermedad suele ser baja y el espectro de enfermos corresponde predominantemente -eso se espera- a casos más precoces y menos severos.

En los programas de tamizaje no se espera que el test aporte un diagnóstico definitivo, pero debe tener sensibilidad y especificidad altas, para no omitir los pocos casos detectables entre el total de sujetos tamizados, y que los falsos positivos que vayan a exámenes confirmatorios sean los menos posibles. Estas condiciones no bastan para considerar efectivo el examen como herramienta de tamizaje. A ellas debemos sumar:

- Que la prueba sea idealmente de bajo costo y complejidad.
- Que el no tratar oportunamente tenga consecuencias clínicamente importantes, y además, que la enfermedad detectada precozmente tenga un tratamiento efectivo.
- Que la prueba sea segura (libre de efectos adversos importantes).
- Que sea aceptable para los profesionales de salud y los pacientes (piense por ejemplo en el tacto rectal como prueba de tamizaje para el cáncer de próstata, para entender a qué se refiere esta condición).

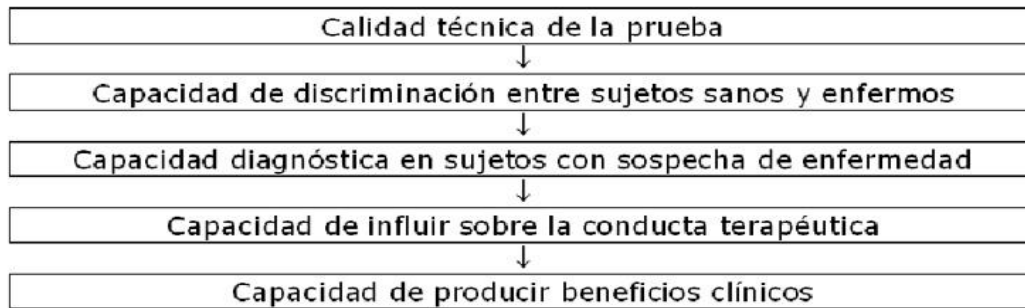
Se puede evaluar este conjunto de factores mediante métodos de simulación (análisis de decisión, estudios de costo-efectividad), pero sin duda la forma óptima de hacerlo es a través de un ensayo aleatorizado en el que una rama del estudio es sometida a tamizaje y la otra no (o a diferentes intensidades de éste), tras lo cual podemos medir el impacto de la intervención.

Por último, los programas de tamizaje también suelen ser evaluados en estudios observacionales (cohortes, caso control), y con distintas formas de estudios de intervención no aleatorizados.

Revisiones sistemáticas de estudios de pruebas diagnósticas

Tal como en otras formas de investigación -por ejemplo, al evaluar la efectividad de una terapia- la forma óptima de sintetizar la evidencia sobre la exactitud de una prueba diagnóstica es mediante una revisión sistemática de la literatura. He aquí un par de ejemplos de este tipo de estudio de revisión:

- P A M Hofman, P Nelemans, G J Kemerink, and J T Wilmink. Value of radiological diagnosis of skull fracture in the management of mild head injury: meta-analysis. J. Neurol. Neurosurg. Psychiatry, April 1, 2000; 68(4): 416-422.
- Vroomen PC, de Krom MC, Knottnerus JA. Diagnostic value of history and physical examination in patients suspected of sciatica due to disc herniation: a systematic review. J Neurol. 1999 Oct;246(10): 899-906.



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-NoComercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.