

Guías para usuarios de la literatura médica

III. Cómo utilizar un artículo sobre un examen diagnóstico

A. ¿Son válidos los resultados del estudio?

Roman Jaeschke, MD, MSc; Gordon H. Guyatt, MD, MSc, y David L Sackett, MD, MSc, por el Evidence-Based Medicine Working Group

(JAMA 1994; 271: 389-391)

ESCENA CLÍNICA

Imagínese que es un médico consultor al que un colega cirujano solicita que visite a una mujer de 78 años, 10 días después de una intervención quirúrgica abdominal, y que presenta una disnea de creciente intensidad durante las últimas 24 horas. La paciente también ha experimentado lo que describe como un malestar torácico que en ocasiones empeora con la respiración profunda (pero otras veces no). Los hallazgos anómalos del examen físico se limitan a un dolorimiento residual en el abdomen y a unas crepitaciones aisladas en ambas bases pulmonares. La radiografía de tórax pone de manifiesto un pequeño derrame pleural derecho, aunque ésta es la primera radiografía desde la intervención. Los gases en sangre arterial indican una PO_2 de 70 mmHg, con una saturación del 92%. En el electrocardiograma, sólo se identifican cambios inespecíficos.

Sospecha que la paciente, a pesar de estar recibiendo 5.000 U de heparina 2 veces al día, puede haber experimentado una embolia pulmonar (EP). Solicita una gammagrafía de ventilación-per-

fusión (gammagrafía V/Q), y el resultado indicado por teléfono a la enfermera es una "probabilidad intermedia" de una EP. Aunque todavía duda del diagnóstico, prescribe una anticoagulación completa. A pesar de que —previamente— ha utilizado a menudo la gammagrafía V/Q y considera que tiene unas nociones bastante buenas de cómo valorar los resultados, se da cuenta de que sus conocimientos se basan en la intuición y en la práctica local antes que en las propiedades de la gammagrafía V/Q a partir de la literatura original. En consecuencia, de camino hacia el departamento de medicina nuclear para revisar la gammagrafía, se detiene en la biblioteca.

LA BÚSQUEDA

Su plan es encontrar un estudio que le indique las propiedades de la gammagrafía V/Q según se aplica a su práctica clínica en general y a esta paciente en especial. Está familiarizado con el programa de software GRATEFUL MED y lo utiliza para su búsqueda. El programa le proporciona una lista de encabezamientos temáticos [*Medical Subject Headings* (MeSH)], y su primera elección es el "embolismo pulmonar". Dado que existen 1.749 artículos con este título MeSH publicados entre 1989 y 1992 (la extensión de su búsqueda), va a tener que reducir la misma. Escoge dos estrategias: elegirá sólo los artículos con "gammagrafía con isótopos radiactivos" como subtítulo y también con el título MeSH asociado "estudio comparativo" (dado que necesita un estudio que compare la gammagrafía V/Q con algún patrón de referencia). Esta búsqueda le depara 31 artículos, de los cuales excluye 11 que evalúan nuevas técnicas diagnósticas, nueve que guardan relación con el diagnóstico y tratamiento de la trombosis venosa profunda y uno que examina la historia natural de la EP. Los 11 artículos restantes tratan de la gammagrafía V/Q en la EP. Sin embargo, uno es un editorial;

cuatro tienen un ámbito limitado (sólo abordan las gammagrafías de perfusión, con situaciones en las cuales la orientación diagnóstica debe empezar con una angiografía pulmonar, o con un defecto de perfusión único). De los restantes, el Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) Study¹ capta su atención, tanto porque está publicado en una revista muy leída, con la que está familiarizado, como porque se hace referencia al mismo en los títulos de varios de los otros artículos. Imprime el resumen de este artículo y comprueba que incluye la siguiente información: entre los individuos con un resultado intermedio de la gammagrafía V/Q, un 33% presentaban una EP. Llega a la conclusión de que ha hecho una buena elección y recupera el artículo de las estanterías de la biblioteca.

Este artículo de la serie "Guías de los usuarios de la literatura médica" y el siguiente presentarán un abordaje para utilizar de forma óptima el artículo.

INTRODUCCIÓN

Los clínicos regularmente se enfrentan a dilemas cuando solicitan e interpretan los exámenes diagnósticos. La incesante proliferación de la tecnología médica hace que sea más importante que nunca la capacidad del clínico para evaluar los artículos sobre exámenes diagnósticos. En consecuencia, este artículo presentará los principios de evaluación eficiente de los artículos sobre exámenes diagnósticos y de utilizar de manera óptima la información que proporcionan. Una vez que decide, tal como se ha ilustrado en el guión clínico con el artículo PIOPED, que un trabajo es potencialmente relevante (es decir, el título y el resumen sugieren que la información es directamente relevante al problema del paciente que está tratando), puede plantear las mismas 3 preguntas que sugerimos en la "Introducción" y en los artículos sobre el tratamiento²⁻⁴ (tabla 1).

De los *Departments of Medicine y Clinical Epidemiology and Biostatistics*, McMaster University, Hamilton, Ontario.

En el primer artículo de esta serie (JAMA 1993; 270: 2.093-2.095) se incluye una lista completa de los miembros (con sus afiliaciones) del Evidence-Based Medicine Working Group. Los siguientes miembros contribuyeron a este artículo: Gordon Guyatt (presidencia), MD, MSc; Eric Bass, MD, MPH; Patrick Brill-Edwards, MD; George Browman, MD, MSc; Deborah Cook, MD, MSc; Michael Farkouh, MD; Hertzell Gerstein, MD, MSc; Brian Haynes, MD, MSc, PhD; Robert Hayward, MD, MPH; Anne Hobbins, MD, PharmD, MSc; Roman Jaeschke, MD, MSc; Elizabeth Juniper, MCSP, MSc; Hui Lee, MD, MSc; Mitchell Levine, MD, MSc; Virginia Moyer, MD, MPH; Jim Nishikawa, MD; Andrew Oxman, MD, MSc, FACPM; Ameen Patel, MD; John Philbrick, MD; W Scott Richardson, MD; Stephane Sauve, MD, MSc; David Sackett, MD, MSc; Jack Sinclair, MD; K.S. Trout, FRCE; Peter Tugwell, MD, MSc; Sean Tunis, MD, MSc; Stephen Walter, PhD, y Mark Wilson, MD, MPH.

Tabla 1. Evaluación y aplicación de los resultados de los estudios de los exámenes diagnósticos

¿Son válidos los resultados del estudio?	
Criterios principales	
¿Existió una comparación ciega, independiente, con un estándar de referencia?	
¿Incluyó la muestra de pacientes un espectro adecuado de pacientes a los cuales en la práctica clínica se aplicará el examen diagnóstico?	
Criterios secundarios	
¿Influyeron los resultados del examen objetivo de la evaluación en la decisión de realizar el estándar de referencia?	
¿Se describieron los métodos para realizar el examen con el suficiente detalle para permitir su reproducción?	
¿Cuáles fueron los resultados?	
¿Se presentan las tasas de probabilidad de los resultados del examen o se proporcionan los datos necesarios para su cálculo?	
¿Me ayudarán los resultados en la asistencia a mis pacientes?	
¿Serán satisfactorias en mi ámbito tanto la reproducibilidad del resultado del examen como su interpretación?	
¿Son aplicables los resultados a mis pacientes?	
¿Modificarán los resultados mi tratamiento?	
¿Se encontrarán mejor los pacientes como consecuencia del examen?	

¿Son válidos los resultados del estudio?

Los métodos utilizados para llevar a cabo un estudio determinan si se pueden creer los resultados del mismo. Afirmar que los resultados son válidos implica que la precisión de los exámenes diagnósticos, según lo descrito, está lo suficientemente cerca de la verdad como para que merezca la pena un examen en profundidad del estudio. En primer lugar, es preciso que determine si puede creer los resultados del estudio considerando la forma como los autores reclutaron a sus pacientes y cómo aplicaron el examen y un estándar de referencia apropiado (u "oro" o "criterio") a los pacientes.

¿Cuáles son los resultados del estudio?

Si decide que los resultados del estudio son válidos, el paso siguiente es determinar la precisión del examen diagnóstico, lo que se lleva a cabo examinando (o calculando por sí mismo) las tasas de probabilidad de los exámenes (a las que a menudo se hace referencia como "propiedades" del examen).

¿Me ayudarán los resultados en la asistencia de mis pacientes?

El tercer paso es decidir la forma de utilizar el examen, tanto para el pacien-

te individual como para la práctica en general. ¿Son generalizables los resultados del estudio?, es decir, ¿puede aplicarlos a este paciente concreto y al tipo de pacientes que visita más a menudo? ¿Cuán a menudo de los resultados del examen resultará probablemente información valiosa? ¿proporciona el examen información adicional más allá de la historia y examen físico? ¿tiene un menor coste o está más fácilmente disponible que otros exámenes diagnósticos para el mismo trastorno objetivo? En último término, ¿se benefician los pacientes si se utiliza este examen?

En este artículo abordamos de manera detallada la primera pregunta, mientras que en el siguiente trabajo de la serie lo haremos con la segunda y tercera preguntas. Para ilustrar el proceso, utilizaremos el artículo PIOPED.

En el estudio PIOPED, 731 pacientes con una sospecha de EP y que dieron su consentimiento se sometieron a una gammagrafía V/Q y a una angiografía pulmonar. Se consideró que la angiografía pulmonar era la mejor forma de demostrar si un paciente presentaba en realidad una EP y, por esta razón, constituía el patrón de referencia. Cada angiografía se interpretó como mostrando uno de 3 resultados: presencia de EP, EP dudosa o ausencia de EP. La precisión de la gammagrafía V/Q se comparó con la angiografía, y los resultados de la gammagrafía V/Q se documentaron en una de 4 categorías: probabilidad elevada (de EP), probabilidad intermedia, probabilidad baja, o casi normal o normal. En las tablas 2 y 3 se muestran las comparaciones de las gammagrafías V/Q y angiografías. Más tarde, abordaremos las diferencias entre estas tablas; por el momento, vamos a aplicar la primera de las 3 preguntas de este artículo.

¿SON VÁLIDOS LOS RESULTADOS DEL ESTUDIO?

Criterios primarios

¿Existió una comparación ciega, independiente, con un patrón de referencia? La precisión de un examen diagnóstico se determina mejor comparándolo con la "verdad". En consecuencia, es preciso que los lectores se aseguren de que a cada paciente se le ha aplicado un patrón apropiado de referencia (tal como una biopsia, cirugía, autopsia o seguimiento a largo plazo), junto con la prueba que se investigaba. En el estudio PIOPED, como estándar de referencia, se utilizó la angiografía pulmonar y hasta tal punto fue un "estándar oro" que pudo lograrse sin

Tabla 2. Relación entre los resultados de las angiografías pulmonares y las gammagrafías de ventilación-perfusión en los pacientes con angiografías satisfactorias

Categoría de la gammagrafía	Angiografía	
	Presencia de embolia pulmonar	Ausencia de embolia pulmonar
Probabilidad elevada	102	14
Probabilidad intermedia	105	217
Probabilidad baja	39	199
Casi normal/normal	5	50
Total	251	480

Tabla 3. Relación entre los resultados de la angiografía pulmonar y la gammagrafía de ventilación-perfusión*

Categoría de la gammagrafía	Angiografía	
	Presencia de embolia pulmonar	Ausencia de embolia pulmonar
Probabilidad elevada	102	14
Probabilidad intermedia	105	217
Probabilidad baja	39	199 + 24
Casi normal/normal	5	50 + 26
Total	251	480 + 150 :

*Incluye a 150 pacientes con una gammagrafía de ventilación-perfusión con una baja probabilidad y resultados casi normal/normal, sin angiografía (136) o no interpretable (14), y sin tromboembolismo clínicamente importante durante el seguimiento.

sacrificar a los pacientes. Al leer los artículos sobre los exámenes diagnósticos, si no puede aceptar el estándar de referencia dentro de lo razonable, (ya que nada es perfecto!), es poco probable que el artículo proporcione resultados válidos para sus objetivos.

Si acepta el estándar de referencia, la siguiente pregunta es si los resultados del examen y el estándar de referencia se evaluaron independientemente entre sí (es decir, por interpretadores que no conocían los resultados de la otra investigación). Nuestra propia experiencia clínica nos indica la razón de que esto sea importante. Una vez que hemos identificado un nódulo pulmonar en una tomografía computarizada, advertimos la lesión, previamente no detectada, en la radiografía de tórax; una vez que conocemos los resultados de la ecocardiografía, auscultamos un soplo cardíaco previamente inaudible. Cuanto más probable es que la interpretación de un nuevo examen pueda estar influida por los conocimientos de los resultados del estándar de referencia (o viceversa), mayor es la importancia de la interpretación independiente de ambos. Los investigadores del PIOPED no indicaron explícitamente en el artículo que los exámenes se interpretarían de manera ciega. Sin embargo, se puede deducir del esfuerzo que hicieron para garantizar lecturas reproducibles, inde-

pendientes, que los interpretadores en realidad eran ciegos, lo que hemos confirmado a través de la correspondencia con uno de los autores. Cuando dichos temas suscitan dudas, la mayor parte de autores están dispuestos a disiparlas si se contacta directamente con ellos.

¿Incluyó la muestra de pacientes un espectro adecuado de los pacientes a quienes se aplicará el examen diagnóstico en la práctica clínica?

Un examen diagnóstico es verdaderamente útil sólo hasta el punto de que distingue entre trastornos o procesos objetivos que de lo contrario podrían confundirse. Casi cualquier examen puede distinguir a un individuo sano de uno gravemente afectado; esta capacidad no nos dice nada sobre la utilidad clínica de un examen. Por consiguiente, el valor pragmático, real, de un examen sólo se establece en un estudio que es lo más parecido posible a la práctica clínica.

Un ejemplo real de cómo las esperanzas suscitadas por la introducción de un examen diagnóstico pueden truncarse a través de las investigaciones posteriores procede de la historia del antígeno carcinoembrionario (ACE) en el cáncer colorrectal. Los valores del antígeno carcinoembrionario, cuando se determinaron en 36 individuos con un cáncer avanzado conocido del colon o recto, eran elevados en 35 de los mismos. Al mismo tiempo, en individuos normales y en una serie de otros procesos se pusieron de manifiesto valores mucho más bajos⁶. Los resultados sugirieron que la determinación de las cifras del ACE podría ser útil en el diagnóstico del cáncer colorrectal o incluso en el cribado de la enfermedad. En estudios posteriores de pacientes con estadios menos avanzados de un cáncer colorrectal (y, por consiguiente, con una gravedad menor del proceso) y pacientes con otros cánceres u otros trastornos gastrointestinales (y, por lo tanto, procesos diferentes, pero potencialmente confundibles), la precisión de las determinaciones ACE cayó en picado y se abandonó la utilización de los valores del mismo para el diagnóstico y cribado del cáncer. En la actualidad, el antígeno carcinoembrionario sólo se recomienda en el seguimiento de los pacientes con un cáncer colorrectal conocido⁷.

En el estudio PIOPED, fueron elegibles y se reclutó a todo el espectro de pacientes con sospecha de EP, incluyendo a los pacientes admitidos con sospecha clínica elevada, media y baja de EP. Por esta razón, podríamos concluir que se escogió una muestra adecuada de pacientes.

Criterios secundarios

Una vez está convencido de que el artículo describe un espectro apropiado de pacientes sometidos a una comparación ciega, independiente, de un examen diagnóstico y un estándar de referencia, muy probablemente sus resultados representan una estimación de la precisión real de la prueba, no sesgada: es decir, una estimación que no distorsiona sistemáticamente la verdad. No obstante, aún se pueden disminuir más las posibilidades de interpretarlo erróneamente considerando otros aspectos.

¿Influyeron los resultados del examen objetivo de la evaluación en la decisión de llevar a cabo el estándar de referencia? Las propiedades de un examen diagnóstico se distorsionarán si sus resultados influyen para que los pacientes se sometan a una confirmación mediante el estándar de referencia. Esta situación, en ocasiones denominada "sesgo de verificación"^{8,9} o "sesgo de la orientación diagnóstica"^{10,11} se aplicaría, por ejemplo, cuando los pacientes con una sospecha de coronariopatía y unas pruebas positivas de ejercicio tuvieran más probabilidades de someterse a una angiografía coronaria (el estándar de referencia) que aquellos con pruebas negativas de ejercicio.

El sesgo de verificación constituyó un problema para el estudio PIOPED; los pacientes, cuyas gammagrafías V/Q se interpretaron como normales o casi normales, y con una baja probabilidad de EP, tuvieron menos posibilidades de someterse a una angiografía pulmonar (69%) que los pacientes con gammagrafías V/Q más positivas (92%). Esto no resulta sorprendente, ya que los clínicos podrían ser reacios a someter a los pacientes con una baja probabilidad de EP a los riesgos de una angiografía. En la tabla 2 se presentan los resultados del estudio PIOPED limitados a los pacientes con una angiografía satisfactoria.

La mayor parte de artículos se detendrían aquí, y los lectores tendrían que concluir que la magnitud del sesgo que es consecuencia de las diferentes proporciones de pacientes con una probabilidad elevada y baja de EP en gammagrafías V/Q, y sometidos a una angiografía adecuada, es incierta, pero probablemente amplia. Sin embargo, los investigadores PIOPED aplicaron un segundo estándar de referencia a los 150 pacientes con una probabilidad baja o gammagrafías normales/casi normales no sometidos a la angiografía (136 pacientes) o, en los cuales, fue dudosa la interpretación angiográfica (14 pacien-

tes); se los consideraría no portadores de una EP si evolucionaban bien sin tratamiento. En consecuencia, fueron seguidos durante un año sin tratamiento con anticoagulantes. Ninguno de estos pacientes desarrolló una EP evidente desde un punto de vista clínico durante este tiempo, a partir de lo cual podemos concluir que en el momento en que se sometieron a la gammagrafía V/Q no existía una EP clínicamente importante (si definimos una EP clínicamente importante como la que requiere una anticoagulación para prevenir acontecimientos adversos posteriores). Cuando estos 150 pacientes, considerados no portadores de una EP mediante este segundo estándar de referencia de un buen pronóstico sin tratamiento anticoagulante, se suman a los 480 con angiografías negativas de la tabla 2, el resultado es la tabla 3. Esperamos que esté de acuerdo con nosotros en que la mejor estimación de la precisión de la gammagrafía V/Q procede de la tabla 3, que incluye a los 150 pacientes, quienes, a partir del seguimiento carecían de una EP clínicamente importante. En consecuencia, utilizaremos estos datos en los cálculos posteriores.

Existieron todavía otros 50 pacientes con gammagrafías de probabilidad elevada o intermedia que no se sometieron a una angiografía o cuyas angiografías no fueron interpretables. Es posible que estos individuos pudieran sesgar los resultados. No obstante, constituyen una proporción relativamente pequeña de la población y, si sus características clínicas no son claramente diferentes de los pacientes sometidos a una angiografía, es poco probable que las propiedades del examen difirieran sistemáticamente en esta subpoblación. Por esta razón, podemos seguir adelante con relativa confianza en los resultados del PIOPED.

¿Se describieron los métodos para llevar cabo el examen con el suficiente detalle como para permitir su reproducción? Si los autores han concluido que se debe utilizar un examen diagnóstico, es preciso que indiquen el modo de utilizarlo. Esta descripción debe abarcar todos los aspectos importantes en la preparación del paciente (dieta, fármacos a evitar, precauciones después del examen), la realización de la prueba (técnica, posibilidad de dolor) y el análisis e interpretación de sus resultados.

Una vez que el lector tiene la seguridad de que los resultados del artículo constituyen una estimación no sesgada de las propiedades del examen, puede determinar con exactitud cuáles son (y en qué grado son útiles) las propieda-

des del examen. Aunque no de una forma impecable (los estudios casi nunca lo son), podemos deducir convincentemente que los resultados constituyen una estimación válida de las propiedades de la gammagrafía V/Q. En el próximo artículo de esta serie describiremos el modo de interpretar y aplicar los resultados. □

Bibliografía

1. The PIOPED Investigators. Value of ventilation/perfusion scan in acute pulmonary embolism: results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED). *JAMA* 1990; 263: 2753-2759.
2. Oxman AD, Sackett DL, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, I: how to get started. *JAMA* 1993; 270: 2093-2095.
3. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. II: how to use an article about therapy or prevention. A: are the results of the study valid? *JAMA* 1993; 270: 2598-2601.
4. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. II: how to use an article about therapy or prevention. B: what were the results and will they help me in caring for my patients? *JAMA* 1994; 271: 59-63.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd ed. Boston, Mass: Litter Brown and Co; 1991: 53-57.
6. Thomson DMP, Krueger J, Freedman SO, Gold P. The radioimmunoassay of circulating carcinoembryonic antigen of the human digestive system. *Proc Natl Acad Sci USA* 1969; 64: 161-167.
7. Bates SE. Clinical applications of serum tumor markers. *Ann Intern Med* 1991; 115: 623-638.
8. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verifications is subject to selection bias. *Biometrics* 1983; 39: 207-215.
9. Gray R, Begg CB, Greenes RA. Construction of receiver operating characteristic curves when disease verification is subject to selection bias. *Med Decis Making* 1984; 4: 151-164.
10. Ranschoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; 299: 926-930.
11. Choid BCK. Sensitivity and specificity of a single diagnostic test in the presence of work-up bias. *J Clin Epidemiol* 1992; 45: 581-586.