

Notas metodológicas

De la evidencia científica a las decisiones clínicas

Examinando la utilidad de una prueba diagnóstica

5/09/2005

Martín Caicoya Gómez-Morán

Servicio de Prevención de Riesgos Laborales del Principado de Asturias. Hospital Monte Naranco.

El director médico del hospital H llamó a su despacho al epidemiólogo clínico para plantearle el siguiente problema: "El número de resonancias magnéticas de rodilla está creciendo de forma exponencial. La lista de espera se alarga, las bajas se prolongan. Todo ello provoca insatisfacción del ciudadano, costes rampantes y descontento del médico por los atrasos. Me gustaría que examinaras la indicación correcta de RMN en patología de partes blandas de rodilla de manera que se pudiera plantear la realización de una guía de práctica clínica."

El epidemiólogo clínico recordó su caso. Le dolía la rodilla y le pidió al traumatólogo que la examinara: "Tienes una lesión meniscal" le dijo tras concluir el examen. "¿Qué hay que hacer?" le preguntó el epidemiólogo. "Operar; de todas formas, hazte una resonancia para quedarnos tranquilos". El epidemiólogo pensó que esa resonancia no tenía objeto si el traumatólogo ya había tomado la decisión de operar. Se preguntó cuántas veces ocurría algo semejante.

El diagnóstico clínico de una lesión meniscal se basa en la sintomatología y el examen físico específico. Hay muchos estudios y varios meta-análisis que evalúan la capacidad clasificatoria de la maniobra diagnóstica de McMurray, la más empleada en la valoración de una posible lesión meniscal. McMurray, un cirujano de Liverpool, basado en su experiencia con 200 meniscos, escribió el artículo clásico¹ en el que decía que con palpación se podían diagnosticar las lesiones anteriores a los ligamentos laterales y que la manipulación "aporta una evidencia válida sobre la existencia de una lesión del segmento posterior de cualquiera de los dos cartílagos."

Desde entonces esta maniobra, llamada de McMurray ha sido empleada como una de las pruebas diagnósticas más fiables. Sin embargo su capacidad clasificatoria está lejos de ser buena. Dependiendo del estudio, es positiva entre el 20 y el 66% de los casos en que hay lesión y lo es también entre el 4 y el 43% de las veces en que no la hay². Por otra parte, la sensibilidad de la palpación de la línea articular se sitúa entre el 58 y el 95% y la especificidad del 5 al 74%. Los autores de esta revisión se quejan en su artículo de la baja calidad de los estudios examinados y concluyen que con los datos manejados tienen que decir que ninguna de las dos pruebas es suficientemente clasificatoria: "... si la existencia de una lesión meniscal tiene consecuencias significativas en el tratamiento, es preciso realizar más estudios".

Sin embargo, varios autores e instituciones consideran que en el contexto de una buena historia clínica, el McMurray y el dolor localizado en la línea articular a la palpación son signos decisivos en el diagnóstico de lesión meniscal^{3,4}

Efectivamente, la potencia diagnóstica de una prueba no sólo depende de su capacidad clasificatoria es decir, de su sensibilidad y especificidad, también depende de la probabilidad que tiene ese paciente de estar enfermo. Es lo que se llama revisión bayesiana de la probabilidad a priori. ¿Es en este caso cierta la afirmación de que el examen físico es más o menos relevante dependiendo de la probabilidad de que el paciente tenga o no la lesión?. Veamos esto en tres supuestos: en el primero la historia clínica es poco sugerente de lesión meniscal, supongamos que 10 de cada 100 pacientes tengan lesión; en el segundo hay una sospecha pero no clara, en un 50% de los pacientes la lesión estará presente y en el tercero el clínico se inclina a pensar que existe una lesión en un 80% de los pacientes. Finalizada la historia clínica, le pide al paciente que se tumben en la camilla para proceder a examinarlo. Aceptemos una sensibilidad del 40% y una especificidad del 70% para el McMurray, cifras que se sitúan en la mitad del rango de las aportadas por Scholten² ¿Cómo interpretar el resultado?

Veamos que ocurriría en cada supuesto:

En el **primer supuesto** la probabilidad de tener la lesión es 100/1000: 10%. La sensibilidad es el cociente entre verdaderos positivos, 40, y el total de enfermos, 100, 40% como habíamos acordado. La especificidad es el cociente entre verdaderos negativos, 630 y el total de no enfermos, 900: 70% como habíamos acordado.

Tabla 1: Primer supuesto

RESULTADO TEST	ENFERMO	NO ENFERMO	TOTAL
POSITIVO	40	270	310
NEGATIVO	60	630	690
TOTAL	100	900	1000

En este caso 310 de las 1000 pruebas realizadas han resultado positivas ¿qué probabilidad hay de que el resultado positivo indique que el paciente está enfermo? La respuesta es fácil: entre los que resultaron positivos a la prueba están enfermos los que lo están: los verdaderos positivos: 40. La probabilidad de que dado un resultado positivo, es decir 310, el paciente esté enfermo, es decir 40 es $40/310$: 13%. El examen físico en un paciente con sospecha baja no añade nada a la sospecha: pasa del 10% al 13%.

Pero más importante es que dado una baja sospecha, ¿El que la prueba fuera negativa me inclina a descartar la patología?. Es el valor predictivo negativo: cociente entre verdaderos negativos y total de negativos. En este caso $630/690$: 93%. El examen físico me deja como antes.

En el **segundo supuesto**: aquí la probabilidad de tener la lesión es del 50%, el valor predictivo positivo (VPP) es $200/350$: 57%, el valor predictivo negativo (VPN) es $350/650$: 54%. El examen físico no aportó nada.

Tabla 2: Segundo supuesto

RESULTADO TEST	ENFERMO	NO ENFERMO	TOTAL
POSITIVO	200	150	350
NEGATIVO	300	350	650
TOTAL	500	500	1000

Finalmente el **tercer supuesto**: la probabilidad de tener la lesión es del 80%, el valor predictivo positivo es $640/760$: 84%, el valor predictivo negativo es $280/1240$: 23%. El examen físico no aportó nada.

La razón de que la prueba no aporte nada es que tiene una baja sensibilidad y no muy buena especificidad. Tras este ejercicio, el epidemiólogo no se atreve a decir, como le hubiera apetecido, que las pruebas rinden mejor en situaciones de alta sospecha clínica. Concluye, con Scholten ² que el examen físico es poco informativo.

Pero una prueba de este tipo no es simplemente positiva o negativa, casi ninguna prueba lo es y menos las que son numéricas: la sospecha de diabetes no es la misma si la glucemia es de 125 que si es de 250. Algo semejante ocurre con la prueba de McMurray: si es fuerte y claramente positiva la sospecha de lesión es mucho más alta que si es débil o dudosamente positiva. En definitiva, hay un juego de sensibilidades y especificidades para cada test en función de la magnitud del resultado. Esta es una de las razones, no la única, por la que la prueba de McMurray tiene una sensibilidad y especificidad tan variable en los diferentes estudios.

Tabla 3: Tercer supuesto

RESULTADO TEST	ENFERMO	NO ENFERMO	TOTAL
POSITIVO	640	120	760
NEGATIVO	960	280	1240
TOTAL	1600	400	2000

Para manejar esa incertidumbre, la derivada de que un test es más informativo según la magnitud de su resultado, es frecuente hoy emplear la razón de verosimilitud, en inglés "likelihood ratio" (LR). La razón de verosimilitudes es el cociente entre ese resultado del test en presencia de enfermedad (sensibilidad) y el de ese resultado en ausencia de enfermedad (1-especificidad o tasa de falsos positivos). Precisamente, la relación entre estas dos variables es la que nos sirve para hacer las curvas ROC, las que se emplea para definir el mejor rendimiento de una prueba.

En las tablas 1, 2 y 3 podemos hallar fácilmente el LR. En la tabla 1 la probabilidad de que el resultado sea positivo entre los enfermos es $40/100$ (que es la sensibilidad) y de que el resultado sea positivo entre los no enfermos es $30/100$ (tasa de falsos positivos). Lo mismo que para las otras dos tablas.

Los LR de los tres casos evaluados es: $40/30=1,33$ La interpretación es que el resultado positivo se encontrará 1,33 más veces en caso de estar enfermo que de no estar enfermo: prácticamente no añade información.

Se puede calcular como varía el LR en función de los diferentes valores posibles de sensibilidad y especificidad. Por ejemplo si la sensibilidad fuera del 66% y la especificidad del 98%, las máximas que aparecen en la tabla 4, el LR+, es decir, la sensibilidad dividido por 1-especificidad sería 33. Sería una

prueba excelente. Como se puede ver en la tabla 4 el LR mayor calculado es 9,5 y no 33. Es porque en nuestro ejemplo se ha forzado la realidad: hay estudios que aportan una sensibilidad del 66%, pero a costa de la especificidad ya que cuanto más bajemos el dintel de aceptación de positivo, más falsos positivos habrá y en consecuencia, menor especificidad. Ellos emplean sensibilidades y especificidades concretas de cada estudio, por eso la máxima es 9,5, un buen LR+.

Tabla 4

Tests	Nº de estudios	Sensibilidad	Especificidad	LR +	LR -
Derrame articular	4	20% a 53%	54% a 100%	0.8 a ∞	0.7 a 1.1
McMurray	11	10% a 66%	57% a 98%	1.5 a 9.5	0.4 a 0.9
Dolor en línea articular	13	28% a 95%	5% a 95%	0.8 a 14.9	0.2 a 2.1
Apley	4	16% a 58%	80% a 99%	0.8 a 19.5	0.5 a 1.1

Finalizado este ejercicio decide comprobar qué recomiendan otros respecto a la RMN. En la única guía de práctica clínica que localizó sobre este tema⁵ lee que la RMN se debe pedir siempre antes de una artroscopia exploradora y cuando la clínica no es suficiente como para diagnosticar la lesión. Siguiendo esa recomendación y la evidencia borrosa de la tabla 4, el epidemiólogo se plantea la posibilidad de hacer un algoritmo diagnóstico en el que se catalogue el resultado del examen físico en varios niveles, por ejemplo, positivo débil con la mayor sensibilidad y menor especificidad, positivo sin cualificar, con sensibilidad y especificidad media y positivo fuerte, con baja sensibilidad y alta especificidad. Suponiendo que haya una sospecha clínica alta, un McMurray fuertemente positivo (sensibilidad 20%, especificidad 95%) tiene un VVP del 94%: no es preciso hacer más estudios. En el otro extremo, una baja sospecha clínica, empleando como prueba un McMurray débilmente positivo (sensibilidad 66%, especificidad 54%) tiene un VPP del 28%, no ayuda, pero su ausencia tiene un VPN del 87%, prácticamente descarta la necesidad de más estudios.

El epidemiólogo, tras examinar estos resultados se plantea definir niveles positividad del McMurray y plantear una guía basada en la sospecha clínica y el grado de positividad del test. Pero considera que la empresa es compleja porque no tiene apoyo empírico para hacerlo. Además sospecha que sería muy difícil porque la prueba es operador dependiente: los más hábiles y expertos realizarán mejor la exploración con lo que el resultado será más informativo, tanto el positivo como el negativo.

También reflexionó sobre la posibilidad de valorar las dos pruebas en conjunto, McMurray y dolor en línea articular. Pero desiste porque aunque sirven ambas como ayuda para diagnosticar lesiones meniscales, con reservas, cada una es más específica de un tipo de lesión meniscal: no son superponibles.

Tomar decisiones es un ejercicio que sólo tiene valor en la incertidumbre. Cuando las instrucciones o la información no permiten más que una opción, su ejecución no se precede de una toma de decisión. Son casos claros en los que sólo hay una forma de abordarlos y la forma de asegurar que se hacen las cosas bien es mediante un procedimiento.

Una guía de práctica clínica, basada en la evidencia, debe contemplar la incertidumbre, además de las preferencias de los pacientes y las circunstancias en las que se realiza el proceso. En este caso, el epidemiólogo no cree que una guía pueda mejorar la solicitud de RMN pues no tiene claro que se pueda definir un algoritmo diagnóstico clínico reproducible que haga innecesaria la confirmación o rechazo mediante resonancia.

De todas formas, el caso no lo considera cerrado. Cree que hay que discutir esas reflexiones con los traumatólogos y radiólogos. Ellos pueden aportar información y opiniones que modifiquen estas conclusiones. Y quizá se deba examinar la práctica clínica revisando un número de historias clínicas. La variabilidad en la práctica, en una patología tan frecuente y con repercusiones tan notables, sería una razón adicional para plantearse la realización de una guía.

El traumatólogo que había pedido una RMN para confirmar el diagnóstico siguió una práctica clínica sancionada por la comunidad científica. Aunque se fiaba de su diagnóstico, sabía que la exploración está sujeta a errores. Manejarse con un abanico de sensibilidades y especificidades en función de la prevalencia, como se ha hecho aquí, es algo que se hace intuitivamente en la clínica. Pero mientras el epidemiólogo se maneja con números, el clínico tiene que hacer frente a enfermedades y errores diagnósticos y cada uno tiene para el enfermo, y para el clínico, una trascendencia imposible de percibir en el número.

Bibliografía

1. McMurray Br . *The semilunar cartilages. Br J Surg* 1942;29:407-14
2. Scholten R.J.P.M., Deville W., Biji D., Opstelten W., van der Plas C.G., Bouter L.M. *The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis. Journal of Family Practice*, 200;50(11):938-44
3. Stratford PW, Binkley J *A review of the McMurray test: definition, interpretation, and clinical usefulness J Orthop Sports Phys Ther.* 1995;22:116-20 .
4. Shoor S. *Review: physical diagnostic tests have low diagnostic accuracy for meniscal lesions of the knee. Evidence-Based Medicine* 2002;7:93
5. *New Zeland Guidelines Group. The diagnosis and management of soft tissue knee injuries: Internal derangements.* (disponible en página web: http://www.nzgg.org.nz/guidelines/0009/ACC_Soft_Tissue_Knee_Injury_Fulltext.pdf)

Palabras clave:

De la evidencia científica a las decisiones clínicas, Razón de verosimilitud

e-notas de evaluación

Autores/as: Martín Caicoya Gómez-Morán

Título: *Examinando la utilidad de una prueba diagnóstica*

Nº 1 de 2005

Disponible en: *Enlace a la noticia*

