¹Unidad de Medicina Basada en Evidencia, Escuela de Medicina, Pontificia Universidad Católica de Chile.

²Departamento de Medicina Familiar.

³Departamento de Medicina Interna, Pontificia Universidad Católica de Chile. Santiago de Chile.

⁴Unidad Docente Asociada, Hospital Dr. Sótero del Río

Correspondencia a:
Dra. Solange Rivera
E-mail: strivera@uc.cl

Aplicabilidad de un estudio sobre tests diagnósticos

SOLANGE RIVERA M.^{1,2}, LUZ MARÍA LETELIER S.^{1,3,4}

Applying results from studies on diagnostic tests

Las decisiones clínicas son el resultado final de un proceso compuesto por una secuencia de acciones en la atención de los pacientes. Se inicia con una adecuada anamnesis y exploración física, y con estos datos se plantean una o más hipótesis diagnósticas. Este proceso puede requerir de exploraciones complementarias (exámenes de laboratorio o imágenes), que ayuden a confirmar o descartar dichas hipótesis. El método clínico se desarrolla, por tanto, a través de la valoración de probabilidades diagnósticas basadas en los resultados de toda la información recogida. Por su importancia en el proceso de toma de decisiones, los tests diagnósticos son elementos esenciales para una práctica clínica de calidad, pero su correcta utilización dependerá de la exactitud de la información que proporcionan¹.

En artículos anteriores, hemos revisado aspectos del proceso diagnóstico², de la validez³ y de la interpretación de los resultados⁴ de un estudio sobre tests diagnósticos. Una vez realizado ese análisis, debemos decidir si esa información es aplicable a nuestros pacientes. El presente artículo tiene por objeto revisar los criterios que debe cumplir un estudio sobre un test diagnóstico para ser aplicado a nuestros pacientes.

Criterios para aplicar los resultados de un estudio sobre tests diagnósticos:

1. ¿Es reproducible el resultado y la interpretación del test en mi medio?

La utilidad de un test dependerá de su capacidad de ser reproducible en la población en la que se pretende aplicar. Esta reproducibilidad puede depender del test en sí mismo (por ejemplo, la variación en los reactivos de un determinado test de

laboratorio) o de su interpretación (por ejemplo, la presencia o tamaño de un nódulo pulmonar en una radiografía). Por esta razón, un estudio que evalúa un test diagnóstico debiera informar la reproducibilidad de sus resultados con alguna medida de concordancia entre los observadores, especialmente en aquellos tests que requieren interpretación o describir en detalle la técnica de laboratorio utilizada.

Por ejemplo: En un estudio sobre un test para el diagnóstico precoz de influenza en niños, en que se comparó un test rápido de influenza con el cultivo o dos PCR consecutivos para influenza A y B⁵, está descrita la manera en que fueron ejecutados cada uno de los test y su correspondiente interpretación. Sin embargo, fue necesario entrenar a los investigadores en este procedimiento. En este caso, independiente de lo útil que pueda ser el test, habría que evaluar la posibilidad real de entrenar a quienes aplicarán el test antes de aplicarlo en la situación real.

2. ¿Son aplicables a mis pacientes los resultados del estudio?

La decisión en este punto es, si el test tendrá la misma exactitud en mi población que la que se informó en el estudio. Esto es importante porque las propiedades de un test pueden variar en diferentes poblaciones de pacientes debido a:

- Diferencias en la severidad de la enfermedad. Un ejemplo es el caso de la ecografía Doppler de extremidades inferiores para el diagnóstico de trombosis venosa profunda (TVP), donde la exactitud del test para el diagnóstico de TVP proximal es mayor en pacientes ambulatorios sintomáticos que en pacientes postoperatorios asintomáticos⁶. En general, en los pacientes

con expresión clara de la enfermedad (ya sea por mayor severidad o tiempo de evolución), el *likelihood ratio* (LR) se aleja más de 1. En cambio en pacientes menos afectados por la condición en estudio (estadios precoces o expresión leve), el LR se acerca al valor 1, es decir, el test tiene menor utilidad.

- b) Diferencias en la distribución de pacientes con enfermedades similares a la condición en estudio⁷. En el caso de poblaciones con enfermedades similares a la condición en estudio, el LR estará muy cercano a 1 y el test parecerá menos útil, por otro lado, si se evalúa en una población con pocos pacientes que cursen con enfermedades similares a la condición en estudio (poca incertidumbre diagnóstica), el LR se alejará de 1 y el test parecerá más útil de lo que realmente es.

En resumen, si el escenario clínico en que uno se desempeña es semejante al escenario en que fue evaluado el test diagnóstico y los pacientes a los que aplicará el test cumplieron los criterios de selección del estudio, es muy probable que los resultados de éste sean aplicables. Si lo anterior no se cumple, es necesario aplicar el juicio clínico e intentar evaluar cuán distintos son mis pacientes a los del estudio en términos de severidad de la enfermedad o de distribución de otras enfermedades semejantes a la condición en estudio como para aplicar los resultados. La introducción de revisiones sistemáticas de estudios sobre tests diagnósticos puede ser de ayuda para resolver este aspecto⁸.

3. ¿Los resultados de este estudio cambiarán mi estrategia de manejo?

Recordando algunos conceptos entregados en artículos anteriores^{2,4}: Para cualquier condición a evaluar hay probabilidades bajo las cuales el clínico podría descartar el diagnóstico y no realizar más tests (umbral de diagnóstico). También habrá un punto sobre el cual la probabilidad de que exista la condición es tan alta que no consideraremos necesario hacer más tests sino iniciar el tratamiento (umbral de tratamiento). Cuando nos encontramos entre estos dos umbrales, será necesario realizar más tests. Para que estos nuevos tests sean de utilidad deben demostrar propiedades discriminatorias que nos permitan cruzar alguno de estos umbrales. Es decir, tests diagnósticos con

LR cercanos a 1, rara vez nos ayudarán a alcanzar estos umbrales. En cambio, un test que tenga LRs muy altos (> 10) o muy bajos ($< 0,1$) moverán considerablemente nuestras probabilidades y posiblemente nos permitirán alcanzar alguno de los umbrales. Por ejemplo, un estudio que midió el rendimiento de un cuestionario breve para demencia, SIS⁹ (Tabla 1), mostró LRs > 10 o $< 0,1$ en 347 de los 651 pacientes estudiados (53%), es decir, este cuestionario cambia la probabilidad post test de más de la mitad de los pacientes en que había sospecha clínica de demencia.

Desde un punto de vista teórico, el test ideal sería aquel que tenga un LR tan alto que permita confirmar la presencia de determinada condición y, a la vez, un LR tan bajo que permita descartar la misma condición. Lamentablemente, en la vida real la mayoría de los tests tienen propiedades adecuadas en un solo sentido, permitiendo confirmar o descartar una condición, pero rara vez permiten lograr ambas.

Por lo anterior, muchas veces es necesario el uso secuencial de distintos tests diagnósticos. Un test puede ser integrado en un flujoograma diagnóstico de las siguientes tres maneras¹⁰:

- Reemplazando a otro test diagnóstico.
- Como *triage* (o selección de pacientes): se realiza previamente y sólo los pacientes con determinado resultado siguen en el proceso diagnóstico.
- Agregándose a un test: se realiza posterior a otro test de tal manera que agrega nueva información, cambiando la probabilidad post-test de la condición en estudio.

Tabla 1. SIS (Six-Item Scores)⁹. Puntajes obtenidos por pacientes y sus respectivos LRs

Puntaje	Demencia +	Demencia -	LR
6	105	2	47
5	64	2	28
4	64	8	7,1
3	45	16	2,5
2	31	35	0,79
1	25	80	0,28
0	11	163	0,06
Total	345	306	

El uso del LR para describir la exactitud de un test se ajusta a estos flujogramas diagnósticos: así, cada elemento de la historia clínica, o hallazgo del examen físico, representa un nuevo test diagnóstico. Podemos utilizar un test para obtener cierta probabilidad post-test que a su vez puede aumentar o disminuir aún más por el siguiente test que realicemos.

Por ejemplo, un estudio¹¹ que evaluó la exactitud diagnóstica de la colonografía por tomografía computada para la detección de cáncer colorectal (CCR) avanzado en pacientes asintomáticos con distintos factores de riesgo, en el subgrupo de pacientes con historia familiar de cáncer colorectal, un examen positivo aumenta a 52% la probabilidad post test de tener CCR, mientras que un examen negativo la disminuye a 2%. Con estas nuevas probabilidades post test, los clínicos pueden tomar la decisión de continuar con el estudio (agregar un nuevo test) o mantener una conducta más expectante.

En suma, antes de aplicar un test diagnóstico debemos decidir si sus propiedades nos permitirán cambiar la probabilidad post test acercándonos a alguno de los umbrales de decisión, lo que cambiaría la estrategia de manejo del enfermo.

4. *¿Los pacientes estarán mejor como resultado de la aplicación del test?*

El último criterio para decidir la aplicabilidad de un estudio sobre un test diagnóstico es responder si el beneficio que conlleva la aplicación del nuevo test supera a los posibles riesgos¹². ¿De qué manera podemos establecer esto? La respuesta está en considerar al test diagnóstico como una maniobra terapéutica. Es decir, no basta que un test tenga buena exactitud diagnóstica. El escenario ideal es que a consecuencia de ese diagnóstico, el paciente obtenga beneficios clínicos derivados del tratamiento oportuno y efectivo; beneficios que no se hubieran obtenido sin la realización del test diagnóstico. Además, es importante que este beneficio supere a los riesgos asociados a la aplicación de este test diagnóstico.

Al igual que en estudios que evalúan una intervención terapéutica, la mejor evidencia para determinar si un test genera más beneficio que daño, son los estudios clínicos aleatorios (ECR). En estos estudios, los pacientes son asignados al azar a distintas estrategias diagnósticas; por ejemplo, a algunos se les aplica un nuevo test y a otros se

aplica el test en uso habitual para luego seguir a los pacientes en el tiempo y evaluar cómo se relaciona el uso de una u otra estrategia a *outcomes* clínicamente relevantes para los pacientes.

¿Cuándo sería de mayor utilidad la información entregada por un ECR de test diagnóstico?

La importancia del uso de un test diagnóstico de gran exactitud es indiscutible si la condición en estudio es dañina para el paciente al no ser diagnosticada, si la aplicación del test tiene un nivel de riesgo aceptable y si existe un tratamiento efectivo para la condición diagnosticada. Este es el caso al usar una angiotomografía axial computada (TAC) de tórax para el diagnóstico de tromboembolismo pulmonar (TEP). Un resultado negativo o un resultado positivo para TEP suele hacer innecesarios nuevos tests o definitivamente confirmar la necesidad de iniciar tratamiento anticoagulante, respectivamente. Otros tests, como el SIS⁹ (cuestionario para demencia), son completamente inocuos, de alto rendimiento y permiten optimizar el manejo del paciente. En estas situaciones no sería necesario contar con un ECR del test diagnóstico.

En cambio, imaginemos un paciente con un carcinoma de pulmón de células no pequeñas, potencialmente resecable; en este caso nos preguntaremos si debemos solicitar una TAC de tórax y según los resultados planificar el manejo o si debemos realizar de inmediato una mediastinoscopia. En este caso se trata de test con buena exactitud diagnóstica, sin embargo, no sabemos si la aplicación de una u otra estrategia diagnóstica podrá cambiar más la mortalidad, los costos o complicaciones del paciente. Para poder contestar esta pregunta, un ECR sería el mejor diseño¹³.

Otro escenario en que son necesarios los ECR de test diagnóstico es el estudio de test de *screening*, donde la aplicación masiva de un test a pacientes sanos debe garantizar no sólo seguridad y exactitud diagnóstica, sino también asegurar que el diagnóstico realizado conlleva la posibilidad de tratamiento y por ende un cambio en el pronóstico de la condición identificada¹⁴. En general, no se discute la exactitud diagnóstica y seguridad de una mamografía o de un antígeno prostático específico (APE), pero su uso masivo en poblaciones sanas debe evaluarse con ECRs que demuestren que su aplicación no sólo identifica correctamente la condición, sino que también demuestran que al tratar la condición detectada por el *screening*, se

disminuye la morbilidad o mortalidad de dicha condición. Una reciente revisión sistemática de ECRs sobre el *screening* con APE muestra un aumento en la detección de cáncer de próstata en estadios precoces, sin embargo, no hubo diferencias en la mortalidad general entre la población con y sin *screening*¹⁵.

Si en el estudio del test rápido de influenza⁵, comentado previamente, el diseño hubiera sido un ECR, la detección precoz de los casos influenza positivos debería generar acciones terapéuticas, como aislamiento e inicio precoz de terapia antiviral, que podrían traducirse en menores complicaciones o reducción de la estadía hospitalaria, *outcomes* que debieran medirse y, según estos resultados, establecer los beneficios del uso del test más allá de su capacidad de diagnosticar la condición.

Conclusión

Cuando se realiza la lectura crítica de un estudio que evalúa la exactitud o la utilidad clínica de un test diagnóstico, los diseños que aportan mayor información al clínico son los estudios de corte transversal y los ensayos clínicos randomizados, respectivamente. La valoración crítica de estudios sobre tests diagnósticos, incluye evaluar el riesgo de sesgo (de acuerdo al diseño), interpretar los resultados y finalmente, decidir si la información es aplicable a nuestra práctica clínica. Para decidir sobre la aplicabilidad, es importante evaluar si el test es reproducible en mi medio, si su capacidad de discriminación será similar en mis pacientes, si el resultado del test cambiará el tratamiento del paciente y finalmente, determinar si el beneficio clínico supera los riesgos y costos que puede traer su aplicación.

Referencias

- Guyatt G, Sackett D, Haynes B. Evaluating diagnostic tests. En: Haynes B, Sackett D, Guyatt G, Tugwell D. *Clinical epidemiology. How to do clinical research*. Lippincott Williams & Wilkins 2006: 273-322.
- Capurro ND, Rada GG. El proceso diagnóstico. Rev Med Chile 2007; 135: 534-8.
- Valenzuela DL, Cifuentes AL. Validez de estudios de test diagnósticos. Rev Med Chile 2008; 136: 401-4.
- Salech F, Mery V, Larrondo F, Rada GG. Estudios que evalúan un test diagnóstico: interpretando sus resultados. Rev Med Chile 2008; 136: 1203-8.
- Poehling K, Griffin M, Dittos R, Tangt Y, Holland K, Li H, et al. Bedside diagnosis of influenza virus infections in hospitalized children. Pediatrics 2002; 110: 83-8.
- Ginsberg JS, Caco CC, Brill-Edwards PA, Panju AA, Bona R, Demers CM, et al. Venous Thrombosis in patients who have undergone major hip or knee surgery: detection with compression US and impedance plethysmography. Radiology 1991; 181: 651-4.
- Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 2002; 137: 598-602.
- Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120: 667-76.
- Callahan CM, Unverzagt FW, Hui SL, Parkins AJ, Hendrie HC. Six-Item Screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002; 40: 771-81.
- Bossuyt PM, Irwing L, Craig J, Glasziou P. Comparative accuracy; assessing new tests against existing diagnostic pathways. BMJ 2006; 332: 1089-92.
- Regge D, Laudi C, Galatola G. Diagnostic Accuracy of Computed Tomographic Colonography for the Detection of Advanced Neoplasia in Individuals at Increased Risk of Colorectal Cancer. JAMA 2009; 301: 2453-61.
- Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A Framework for clinical evaluation of diagnostic Technologies. CMAJ 1986; 134: 587-94.
- Canadian Lung Oncology Group. Investigation for mediastinal disease in patients with apparently operable lung cancer. Ann Thorac Surg 1995; 60: 1382-9.
- Newman T, Browner W, Cummings S, Hulley S. Designing Studies of medical tests. En: Hulley S, Cummings S, Browner W, Grady D, Newman T, (eds). Designing Clinical research. Lippincott Williams & Wilkins 2006; 273-322.
- Djulbegovic M, Beyth RJ, Neuberger MM, Stoffs TL, Vieweg J, Djulbegovic B, et al. Screening for prostate cancer: systematic review and meta-analysis of randomized controlled trials. BMJ 2010; 341: c4543. doi: 10.1136/bmj.c4543.