

¿Cómo analizar artículos sobre tests diagnósticos en medicina intensiva? El caso del fracaso de la extubación debido a edema laríngeo

R. FERNÁNDEZ FERNÁNDEZ

Centro de Críticos. Hospital de Sabadell. Corporación Sanitaria–Instituto Universitario Parc Taulí. Sabadell. Barcelona. España.

INTRODUCCIÓN

La atención médica directa es la base de la actividad mayoritaria en medicina intensiva. Para ello utilizamos diariamente herramientas diagnósticas sobre las que articular diagnósticos, ya sean sindrómicos o fisiopatológicos. Dichos tests diagnósticos pueden proceder de la adaptación de los valores obtenidos como normales en poblaciones sanas o, más frecuentemente, de poblaciones de referencia hospitalarias.

Debido a las graves alteraciones de órganos y sistemas que origina el ingreso en la UCI, los valores a partir de los que diagnosticamos la significación clínica de una anomalía en nuestros pacientes suelen ser muy distintos de la “normalidad”.

El término normalidad puede tener dos acepciones claramente distintas: la normalidad fisiológica y la estadística¹. Como normalidad fisiológica definimos aquellos parámetros que definen el estado correcto u óptimo de una función. De esta forma, es normal tener una temperatura corporal de 36 °C en reposo. Pero también es normal tener una temperatura de 38 °C durante una infección gripal. Generalmente este concepto de normalidad se asocia a la idea de que es algo bueno o beneficioso para el paciente. Cuando hablamos de normalidad como función estadística nos referimos al valor que probabilísticamente vamos a encontrar con más frecuencia cuando observemos a pacientes en determinada

condición, independientemente del juicio de valor sobre si ello es beneficioso o no. Llevado al extremo, podríamos concluir que la presión arterial sistólica estadísticamente normal en pacientes con infecciones muy graves en el momento del ingreso en la UCI es de 90 mmHg. En este caso es evidente que el concepto de normalidad estadística puede ser completamente diferente del concepto de normalidad fisiológica o de “bondad”.

Tras estos preliminares, emplearemos el análisis de un artículo, recientemente publicado en una de las mejores revistas internacionales de medicina intensiva, como modelo para revisar los aspectos que se deben evaluar cuando se habla de tests diagnósticos. El trabajo de Yann de Bast et al² evalúa la utilidad del test de la fuga del neumotaponamiento del tubo endotraqueal para detectar qué pacientes fracasarán en la extubación debido a un edema laríngeo. Los autores estudian a 76 pacientes intubados durante más de 12 h, de los cuales el 11% requirió reintubación por edema laríngeo. Los pacientes que requirieron intubación habían presentado una fuga de aire menor que los otros (el 9 frente al 35% del Vt insuflado por el ventilador). El mejor punto de corte con valor diagnóstico fue una fuga del 15,5%. Los pacientes con fuga aérea elevada (definida como > 15,5%) requirieron reintubación en el 3% de los casos, comparado con el 24% en el grupo de fuga aérea baja. La sensibilidad del test fue del 75%; la especificidad, del 72,1%; el valor predictivo positivo (VPP), del 25%; el valor predictivo negativo (VPN), del 96,1%, y se clasificó correctamente (*accuracy* en inglés) el 72,4%. Los autores concluyen que la fuga aérea alrededor del tubo endotraqueal superior al 15,5% puede usarse como cribado para limitar el riesgo de reintubación por edema laríngeo.

Correspondencia: Dr. R. Fernández.
Centro de Críticos. Hospital de Sabadell.
Parc Taulí, s/n. 08208 Sabadell. Barcelona. España.
Correo electrónico: rfernandez@cspt.es

Manuscrito aceptado 17-III-2003.

Razonablemente es de esperar que, tras la lectura de un artículo publicado en una revista de prestigio y con revisión por iguales, los médicos pudiéramos suscribir las conclusiones de los autores sin tener que hacer ningún esfuerzo crítico adicional. El objetivo del presente trabajo es recordar cuán ilusoria es tal presunción y cómo ayudar a los lectores a juzgar más correctamente la bibliografía biomédica.

Cuando evaluamos críticamente un estudio sobre tests diagnósticos deberíamos ser capaces de contestar a una serie de preguntas sobre la validez intrínseca del estudio para, posteriormente, plantear si los resultados además de válidos son extrapolables a un paciente concreto o, por lo menos, al grueso de nuestra actividad clínica³.

De forma esquemática se aplicarán estas cuestiones extendidas a través de la medicina basada en la evidencia al artículo en cuestión.

¿SON VÁLIDOS LOS RESULTADOS DEL ESTUDIO?

Para ello hay que analizar hasta qué punto los autores seleccionaron adecuadamente la muestra a estudio (como aproximación a la totalidad de la población en riesgo); si no se desviaron o sesgaron en la aplicación del test que se somete a estudio y si analizaron correctamente el resultado desde el punto de vista clínicamente relevante.

¿Hubo una comparación ciega e independiente con un estándar de referencia?

En este caso el concepto de comparación ciega requeriría que quien diagnosticara el edema laríngeo no conociera el resultado del test de fuga aérea. El concepto de independiente supone que la realización de la laringoscopia no dependa del resultado del test de fuga aérea.

Los autores refieren que los resultados del test de la fuga aérea no fueron dados a conocer a los médicos que atendían a los pacientes y, por tanto, no podían verse influidos en el diagnóstico de referencia del edema laríngeo. Es interesante destacar que el test no pretendía diagnosticar el edema laríngeo, sino sólo aquel de suficiente gravedad como para requerir reintubación. En este punto, el diagnóstico de edema laríngeo se confirmaba por laringoscopia en el momento de reintubar, pudiendo separar claramente a los pacientes que habían fracasado sólo por edema laríngeo y lo que lo habían hecho por otras causas. Los autores especifican que “los pacientes que requerían intubación por otros motivos, como retención de secreciones bronquiales o deterioro del estado general, no se consideraba que tuvieran edema laríngeo”. No obstante, no parece descabellado pensar que algunos de los pacientes con edema laríngeo también tuvieran broncopneumopatía u otras complicaciones.

¿Incluyó la muestra de pacientes una variedad similar a la que se aplicaría el test en la práctica?

En este punto ya encontramos la primera discrepancia con la mayoría de estudios en este campo. Los autores incluyen a pacientes en postoperatorio reciente (< 48 h) y a otros con intubación más prolongada. Es conocido que los pacientes ventilados durante < 48 h tienen una incidencia muy baja de problemas postextubación, como confirma este estudio, con una incidencia de reintubación del 5%, y ninguno de ellos por edema laríngeo. Atendiendo a ello, sería más fácil extraer conclusiones sólo a partir del grupo de pacientes intubados por > 48 h.

¿Influyó el resultado del test a estudio en la determinación o no del estándar de referencia?

Como ya se ha comentado, parece que el enmascaramiento del test fue suficiente para no influir en el diagnóstico de fracaso de la extubación por edema laríngeo.

¿CUÁLES SON LOS RESULTADOS?

Una vez conocidas las limitaciones en cuanto al impacto de la selección de la población en la validez del test, deberíamos analizar cuál es el poder diagnóstico real del test.

¿Se han presentado las *likelihood ratios*, sensibilidad y especificidad, o por lo menos se dispone de los datos para poder calcularlas?

Los autores aportan la sensibilidad, la especificidad, el VPP y el VPN del test para la población total estudiada. Incluso plantean cómo podrían cambiar tales parámetros dependiendo del punto de corte elegido según el objetivo clínico. Si el objetivo clínico fuera reducir el riesgo de mantener innecesariamente la intubación prolongada podríamos usar un test de fuga con el punto de corte en sólo un 1%, pero entonces la sensibilidad caería al 25%, es decir, sólo detectaríamos la cuarta parte de pacientes en que la extubación fracasara por edema laríngeo. Si nuestro objetivo fuera minimizar el riesgo de fracaso de la extubación, podríamos elevar el punto de corte hasta el 23% de fuga, pero entonces el VPP sería sólo del 47%, es decir, más de la mitad de los pacientes serían incorrectamente diagnosticados de edema laríngeo, sin padecerlo, y probablemente no fueran extubados. El punto de equilibrio de igualar unos y otros lo aporta matemáticamente la curva ROC en el citado punto de corte del 15,5%.

Otra forma de expresar los resultados que aporta otro tipo de información son las *likelihood ratios* (LR), que tienen en cuenta, además, la probabilidad existente de que se produzca el desenlace clínico y cómo ello afecta a la fiabilidad del test. Los parámetros habituales de VPP y VPN son altamente dependientes de la frecuencia con que se presenta la enfer-

TABLA 1. Simulación en la que un test con la misma sensibilidad y especificidad muestra como cambia el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN) cuando la frecuencia de la enfermedad es del 50% (panel A) o de sólo el 9% (panel B); sin embargo, las *likelihood ratio*, tanto para un valor positivo (4,00) como negativo (0,12), se mantienen en ambos casos

Panel A	Enfermedad +	Enfermedad -	Sensibilidad	Especificidad	VPP	VPN
Test +	20	5	91%	77%	80%	89%
Test -	2	17				

Panel B	Enfermedad +	Enfermedad -	Sensibilidad	Especificidad	VPP	VPN
Test +	20	50	91%	77%	29%	99%
Test -	2	170				

medad en estudio. En la tabla 1 se muestra una simulación en la que un test con las mismas sensibilidad y especificidad muestra grandes cambios en VPP y VPN cuando la prevalencia de la enfermedad pasa del 50 al 9%. Con los datos del estudio, se pueden calcular en forma de tabla de 2 x 2 las posibilidades de que un test positivo o negativo se asocie con edema laríngeo o no (tabla 2). Igualmente se pueden calcular estos parámetros sólo en el grupo de intubados > 48 h, probablemente con más significado clínico, como ya se ha explicado (tabla 3).

Con estos datos, la conclusión es que, en los enfermos intubados > 48 h, un test de fuga aérea < 15,5% es 2,25 veces más probable en pacientes que requerirán reintubación por edema laríngeo que en los que no lo presentarán. Por el contrario, un test de fuga aérea ≥ 15,5% es tres veces (la inversa de 0,38) más probable en pacientes que no precisen reintubación por edema laríngeo. Para ponerlo en

perspectiva, se suele requerir una LR(+) > 5 y LR(-) < 0,10 para que un test se considere realmente útil para tomar decisiones clínicas.

¿ME AYUDARÁN LOS RESULTADOS A CUIDAR A MIS PACIENTES?

¿Se describen los métodos suficientemente para ser replicados?

La descripción de la técnica de la fuga aérea sí es suficiente para que cualquier lector pueda llevarla a cabo en sus pacientes sin problemas.

¿Serán satisfactorias la reproducibilidad del test y su interpretación en mi ámbito?

En principio, los resultados de este estudio habrían de poderse reproducir en cualquier UCI medicointensiva estándar, teniendo en cuenta la descripción de los medios y el perfil de los pacientes.

¿Son aplicables los resultados a mis pacientes?

En este punto es donde más frecuentemente los clínicos erramos al trasladar directamente los resultados de los estudios a nuestros pacientes, sin evaluar exactamente en qué se parecen los pacientes estudiados a los nuestros. Generalmente intentamos la comparación contrastando las características clínicas de los pacientes en la inclusión en el estudio. Según esto, la mayoría de los pacientes de una UCI media parecerían comparables en cuanto a gravedad (SAPSII 24-41), volumen corriente, PEEP y diámetro del tubo endotraqueal. No obstante, para poder saber si estos pacientes se parecen a los nuestros debemos comparar otros aspectos, principalmente la evolución, y es aquí donde este trabajo presenta mayores dificultades. Es destacable que en pacientes con intubación > 48 h el índice de reintubaciones sea del 30% (12/38) cuando en la mayoría de series recientes oscila entre el 15 y el 20%. Pero todavía más sorprendente es que en ellos la principal causa de reintubación sea el edema laríngeo, con un 66%

TABLA 2. Comparación de los resultados del test de fuga aérea para diagnosticar fracaso de la extubación por edema laríngeo en la población total

	Edema laríngeo	Ausencia de edema laríngeo
Fuga aérea < 15,5%	6	19
Fuga aérea ≥ 15,5%	2	49
Total	8	68

Sensibilidad = (6/8) x 100 = 75%.
Especificidad = (49/68) x 100 = 72,1%.
Likelihood ratio para un test positivo = Sensibilidad/(1 - Especificidad) = 2,68.
Likelihood ratio para un test negativo = (1 - Sensibilidad)/Especificidad = 0,35.

TABLA 3. Comparación de los resultados del test de fuga aérea para diagnosticar fracaso de la extubación por edema laríngeo sólo en intubados > 48 h

	Edema laríngeo	Ausencia de edema laríngeo
Fuga aérea < 15,5%	6	10
Fuga aérea ≥ 15,5%	2	20
Total	8	30

Sensibilidad = (6/8) x 100 = 75%.
Especificidad = (20/30) x 100 = 66,7%.
Likelihood ratio para un test positivo = Sensibilidad/(1 - Especificidad) = 2,25.
Likelihood ratio para un test negativo = (1 - Sensibilidad)/Especificidad = 0,38.

(8/12), cuando los propios autores reconocen que el promedio suele ser entre el 15 y el 19%. Sólo en una serie de pacientes seleccionados con traumatismo la incidencia fue superior (38%) pero todavía menos de la mitad que en el presente estudio. Tampoco parece un perfil habitual de pacientes que el 80% de los enfermos que presentó estridor precisara reintubación, sugiriendo una casi nula respuesta al tratamiento. Aunque no dispongamos de la respuesta que aclare por qué los pacientes de este estudio presentaron tanta incidencia de edema laríngeo, esto impide extrapolar los resultados a nuestra población si en el problema clínico que se examina hay tantas diferencias.

¿Pueden estos resultados cambiar mi tratamiento?

Incluso aceptando la conclusión más prudente del estudio, es decir, que los pacientes con una fuga aérea muy elevada están a salvo de presentar edema laríngeo, no podemos dejar de tenerlos vigilados, puesto que siguen presentando riesgo de precisar reintubación por otros motivos. Tampoco los pacientes con fuga aérea pequeña verán alterado su tratamiento, dado que la mayoría no sufrirá edema laríngeo y no sería razonable no extubarlos. Únicamente si existiera una alternativa terapéutica entre el test y el desarrollo de fracaso ventilatorio podríamos sugerir tratarlos a todos postextubación, siempre y cuando los riesgos asociados a dicho tratamiento fueran mínimos.

¿Evolucionarán mejor mis pacientes con el test?

Mientras no se prevea un cambio en el tratamiento no es anticipable que la evolución de nuestros pacientes vaya a ser mejor con la inclusión del test de fuga aérea en la atención rutinaria de los pacientes sometidos a intubación endotraqueal.

En resumen, cabe recordar que al enfrentarnos a un artículo sobre tests diagnósticos en medicina intensiva debemos evaluar no sólo su validez metodológica como herramienta, sino también su aplicabilidad en tanto que las poblaciones de pacientes o la experiencia de los investigadores puedan ser claramente comparables con nuestra realidad clínica.

Nota aclaratoria final

Como expresión de las posibilidades de existir errores incluso tras revisiones por pares exquistas, se ha detectado un error en el artículo original. En él se cita la inclusión de 76 pacientes, cuando la suma total en el diagrama muestra 77 episodios. La diferencia radica en el grupo de ventilación > 48 h y nivel bajo de fuga.

BIBLIOGRAFÍA

1. Sanchis Aldas J. La necesidad de objetivar en medicina y sus tribulaciones. Arch Bronconeumol 2002;38:431-5.
2. De Bast Y, de Backer D, Moraine JJ, Lemaire M, Vanderborgh C, Vincent JL. The cuff leak test to predict failure of tracheal extubation for laryngeal edema. Intensive Care Med 2002; 28:1267-72.
3. Jaeschke RZ, Meade MO, Guyatt GH, Keenan SP, Cook DJ. How to use diagnostic test articles in the intensive care unit: diagnosing weanability using f/Vt. Crit Care Med 1997;25: 1514-21.