

Lectura crítica de artículos

C. Ochoa Sangrador^a, P. Martín Muñoz^b,
J. Ruiz-Canela Cáceres^c, J.C. Buñuel Álvarez^d

^aServicio de Pediatría. Hospital Virgen de la Concha. Zamora.

^bSecretaría de Calidad. Hospital Universitario Virgen del Rocío. Sevilla.

^cDistrito Sanitario Sevilla. Sevilla.

^dÁrea Básica de Salud Girona-4. Girona.

Rev Pediatr Aten Primaria. 2007;9 Supl 2:S151-60

Carlos Ochoa Sangrador, cochoas@meditex.es

Resumen

La medicina basada en la evidencia nos ofrece herramientas de gran utilidad para poder resolver problemas clínicos mediante el análisis eficiente de la literatura científica. Herramientas metodológicas que, si adquirimos y ejercitamos, nos ayudarán a valorar cualquier evidencia científica y a integrarla con nuestros conocimientos y experiencia clínica para poder decidir sobre su aplicabilidad e idoneidad en un paciente concreto.

En esta exposición repasaremos los principios generales de la valoración crítica de la literatura científica. También revisaremos los principales criterios que se deben considerar en la valoración de la validez y aplicabilidad de los estudios de evaluación de intervenciones sanitarias y de pruebas diagnósticas.

Palabras claves: Medicina basada en la evidencia, Ensayo clínico controlado aleatorio, Publicaciones, Guías.

Abstract

The evidence based medicine offers us very useful tools to solve clinical problems by means of the efficient analysis of scientific literature. If we acquire and exercise these methodological tools, we will be able to value any scientific evidence and integrate it with our knowledge and clinical experience, and also we will be able to decide on its applicability and suitability in a concrete patient.

In this article we will review the general principles of the critical appraisal of scientific literature. Also we will review the main criteria to consider in the analysis of the validity and applicability of the studies of evaluation of health interventions and diagnostic tests.

Key words: Evidence-based medicine, Randomized controlled trial, Publications, Guidelines.

Introducción

En nuestro ejercicio profesional nos enfrentamos a menudo a situaciones en las

que se nos plantean dudas sobre distintos aspectos de la práctica clínica: etiología, diagnóstico, tratamiento, etc. Tradicional-

mente hemos intentado solucionar nuestras dudas a través de consultas a libros de texto, consultas a revistas o preguntando a colegas. Pero este abordaje clásico presenta importantes limitaciones. Como alternativa, la medicina basada en la evidencia (MBE) propone un método estructurado para resolver dudas clínicas que comprende cuatro pasos¹:

1. Convertir nuestra duda en una pregunta clínica estructurada.
2. Realizar una búsqueda bibliográfica para encontrar artículos que puedan responderla.
3. Realizar una valoración crítica de los documentos recuperados analizando la validez y relevancia de los resultados.
4. Integrar la valoración realizada con nuestra experiencia clínica, considerando la aplicabilidad de los resultados en nuestros pacientes, y actuar.

El objetivo de la valoración crítica es analizar la validez y la aplicabilidad de las evidencias publicadas, para lo que se requieren ciertos conocimientos y habilidades que es preciso adquirir y ejercitar²⁻⁵.

Evaluación de una intervención sanitaria

¿Son válidos sus resultados?

Dado que una intervención sanitaria, ya sea terapéutica o preventiva, es un

factor de estudio susceptible de ser controlado por el investigador, el diseño que proporciona las mejores pruebas sobre su eficacia es el ensayo clínico aleatorizado (ECA) (tabla I). La característica metodológica clave del ECA es la asignación aleatoria de los sujetos a los grupos de estudio ya que, al intervenir solamente el azar, tiende a asegurar que se produce una distribución equilibrada de todas las variables, tanto conocidas como desconocidas, entre los grupos. Para que ni el sujeto ni el investigador puedan influir en la decisión de qué intervención recibirá cada uno de los participantes, es igualmente importante que se aplique alguna técnica de enmascaramiento en el proceso de asignación.

Otro aspecto importante que se debe considerar es el seguimiento completo de los grupos de estudio. La pérdida en el seguimiento de un número elevado de sujetos puede modificar los resultados finales, ya que su salida del estudio puede estar relacionada con una evolución diferente, que en ocasiones habrá sido más desfavorable y en otras más favorable que la del resto de la muestra.

Al igual que en la práctica clínica, en los ECA muchos sujetos incumplen la intervención prescrita o no se les puede aplicar por diversas circunstancias. En esta situación puede constituir un error excluir a

Tabla I. Criterios de evaluación de una intervención sanitaria
Validez de los resultados
<ul style="list-style-type: none"> - Asignación aleatoria - Enmascaramiento de la asignación - Seguimiento completo - Análisis por intención de tratar/por protocolo - Diseño ciego - Homogeneidad inicial de los grupos de estudio - Manejo homogéneo de los grupos de estudio
Magnitud de los resultados
<ul style="list-style-type: none"> - Medidas de efecto (RRR, RAR, RR, OR, diferencias de medias) - Control de variables de confusión - Intervalos de confianza
Aplicabilidad de los resultados
<ul style="list-style-type: none"> - Adecuación del ámbito de estudio y criterios de inclusión-exclusión - Relevancia clínica de la variable principal y las secundarias - Repercusión clínica de los resultados (NNT) - Adecuación del análisis de subgrupos - Valoración de perjuicios y costes
<p><i>RRR: reducción relativa del riesgo; RAR: reducción absoluta del riesgo; RR: riesgo relativo; OR: odds ratio; NNT: número necesario a tratar.</i></p>

esos sujetos del análisis, ya que las razones por las que un tratamiento no se realiza tienen implicaciones pronósticas, y la exclusión introduciría un sesgo en el estudio. En el análisis por intención de tratar, los datos se analizan considerando a cada sujeto como si hubiera recibido la intervención que se le asignó inicialmente y no la que realmente recibió (análisis por protocolo). Esta estrategia, controvertida, pretende controlar posibles sesgos, tanto

conocidos como desconocidos, aunque no siempre es apropiada.

El conocimiento por parte de los pacientes del grupo de tratamiento al que pertenecen tiende a modificar la opinión sobre su eficacia. Igualmente ocurre con los investigadores, en quienes puede inducir un comportamiento diferenciado en el manejo o la evaluación de la respuesta. La mejor manera de evitar posibles sesgos es mantener, en la medida de lo posible,

el ciego respecto al tratamiento tanto del paciente (simple ciego) como de los investigadores (doble ciego).

Tal y como mencionamos anteriormente, la asignación aleatoria de un ECA pretende garantizar la comparabilidad de los grupos al inicio del estudio, de modo que lo ideal sería que sólo se diferenciara en la intervención que van a recibir. No obstante, resulta fundamental realizar una comprobación de que los grupos obtenidos han sido finalmente homogéneos, especialmente si las muestras son pequeñas.

Asimismo, es importante comprobar que el manejo y el seguimiento de los distintos grupos han sido homogéneos a lo largo del estudio. Cualquier otra intervención aplicada a los sujetos de estudio y que pudiera influir en los resultados debería ser equiparable en ambos grupos o controlada en el análisis.

¿Cuáles son los resultados?

Con frecuencia los ECA utilizan como criterio de medición de los resultados de una intervención la presencia o ausencia de un evento adverso o favorable. En el caso de una intervención beneficiosa, su efecto debería reflejarse en la existencia de una menor proporción de eventos en el grupo de intervención que en el grupo control. La forma más simple de es-

timar ese efecto es calcular la diferencia absoluta entre ambas proporciones, lo que se conoce como reducción absoluta del riesgo (RAR). Otra medida de efecto, comúnmente referida en las publicaciones, es la reducción relativa del riesgo (RRR), que ajusta la RAR a una escala relativa y que expresa la reducción del riesgo respecto al riesgo en el grupo control.

Todas estas medidas vendrán referidas o podrán calcularse a partir de los resultados contenidos en el trabajo examinado. Pero a pesar de que son las mejores medidas disponibles, nunca podremos estar absolutamente seguros de que reflejan la verdadera reducción del riesgo, ya que son estimaciones puntuales obtenidas a partir de muestras. No obstante, es previsible que el verdadero valor se encuentre en las proximidades de esta estimación, tanto más próximo cuanto más grande sea la muestra; se puede concretar esa proximidad en el intervalo de confianza.

El intervalo de confianza contiene información de interpretación más intuitiva que el clásico nivel de significación estadística ($p < 0,05$), ya que el lector puede juzgar cuáles son las estimaciones más favorables y desfavorables de la reducción del riesgo (límites superior e inferior del intervalo). No obstante, tanto el cál-

culo de los intervalos de confianza como del nivel de significación no deben sustraer al lector de la consideración de la significación clínica de los resultados obtenidos.

No todos los ECA utilizan como criterio de medición variables dicotómicas. En ocasiones, el efecto se refiere a una variable cuantitativa. En estos casos, la medida del efecto de la intervención podrá estimarse calculando la diferencia entre los valores medios de dichas variables en los distintos grupos de estudio (diferencia de medias o medianas).

¿Son aplicables en tu medio?

A la hora de juzgar si los resultados del estudio son aplicables a nuestros pacientes es preciso tener en cuenta la manera en que nuestro entorno de trabajo se parece al ámbito en el que los pacientes han sido reclutados, si cumplen o no los criterios de inclusión y exclusión del estudio y si tienen condiciones de gravedad o comorbilidad diferentes que puedan interferir en el resultado.

En ocasiones, los autores del trabajo presentan en los resultados información diferenciada de subgrupos de pacientes, en los que las medidas de efecto son más o menos favorables. Este hecho requiere especial precaución cuando los resultados globales son menos favora-

bles. Podríamos considerar los resultados del análisis de subgrupos si la diferencia del efecto es grande, no atribuible al azar, el análisis ha sido planificado a priori y resulta concordante con los resultados de otros estudios.

Para poder planificar la aplicabilidad del estudio es preciso valorar la importancia clínica del parámetro empleado en la medición de los resultados. Existen algunos parámetros (ejemplo: supervivencia) que no ofrecen dudas en cuanto a su repercusión clínica; sin embargo, existen otros (por ejemplo: escalas subjetivas de valoración de síntomas, indicación de hospitalización), cuyo significado clínico es más problemático.

Otro aspecto importante a la hora de juzgar la significación clínica de un resultado es considerar también los posibles efectos adversos no deseados ligados a la intervención aplicada. En la decisión final sobre la aplicabilidad de los resultados tendrán que valorarse la magnitud y la importancia tanto de los beneficios como de los perjuicios.

La magnitud del efecto esperado en un paciente puede estimarse a partir de las medidas de RAR y RRR. Sin embargo, existe otra medida que ofrece una información más intuitiva de los posibles beneficios o perjuicios: el número necesario a tratar (NNT). El NNT es el número de

sujetos que se necesitaría tratar con una intervención específica para producir, o evitar, un evento determinado. El equivalente al NNT para la aparición de efectos adversos no deseados se denomina número necesario a perjudicar (NNP). El NNT y su intervalo de confianza pueden calcularse fácilmente a partir de los inversos de la RAR y los límites de su intervalo de confianza.

Finalmente, también debemos tener en cuenta la repercusión clínica de los resultados y el coste, la facilidad de aplicación, el grado de cumplimiento y la accesibilidad de la intervención considerada.

Evaluación de una prueba diagnóstica

¿Son válidos sus resultados?

Para poder valorar la validez de una prueba diagnóstica es preciso comparar sus resultados con los de un patrón de referencia en una serie de pacientes (tabla II). El patrón de referencia empleado tiene que contar con una validez contrastada o, al menos, aceptada por consenso. La utilización de un patrón de referencia defectuoso puede introducir sesgos en las estimaciones de validez de la prueba.

En relación con el patrón de referencia, resulta también importante considerar si es capaz de clasificar el estado

de enfermedad en todas las observaciones. En el caso de que existan observaciones con un diagnóstico indeterminado, si éstas se excluyen del análisis, se producirán estimaciones sesgadas de las características operativas de la prueba diagnóstica. Este sesgo, conocido como sesgo por exclusión de indeterminados, ocasiona habitualmente sobrestimaciones de la sensibilidad y de la especificidad.

Si aceptamos la validez del patrón de referencia, la siguiente cuestión que hay que tener en cuenta es si tanto la prueba diagnóstica como el patrón de referencia han sido realizados de forma independiente. Cuando no se realizan de forma independiente, puede existir un sesgo de revisión si el resultado de una prueba es susceptible de interpretación subjetiva y se ve influido por el conocimiento del diagnóstico o de las características clínicas del paciente. Para poder garantizar la validez de las estimaciones, deberían realizarse de forma ciega la prueba diagnóstica y el patrón de referencia.

Otro aspecto importante que condiciona la validez del estudio es la inclusión en éste de un adecuado espectro de pacientes, similar al que nos encontramos en nuestra práctica clínica. Para valorar esta cuestión es preciso que los

Tabla II. Criterios de evaluación de una prueba diagnóstica
Validez de los resultados
<ul style="list-style-type: none"> - Patrón de referencia válido - Comparación ciega de la prueba diagnóstica y el patrón de referencia - Espectro de pacientes adecuado - Descripción adecuada de los métodos
Magnitud de los resultados
<ul style="list-style-type: none"> - Presentación correcta de resultados - Cocientes de probabilidades - Análisis de probabilidades antes y después de la prueba - Curvas ROC - Intervalos de confianza
Aplicabilidad de los resultados
<ul style="list-style-type: none"> - Validez externa de la reproducibilidad e interpretación - Adecuación del espectro de pacientes - Repercusión sobre el manejo diagnóstico/terapéutico del paciente - Beneficio sobre el paciente del resultado de la prueba

criterios de selección y las características clínicas y epidemiológicas de la muestra analizada estén claramente presentados.

El diseño del estudio debe tratar de garantizar que en la muestra no se hayan excluido pacientes en función del resultado de la prueba o de la existencia de mayor o menor riesgo de enfermedad. Incurriremos en un sesgo de verificación diagnóstica cuando la probabilidad de que se les realice el patrón de referencia sea menor entre los sujetos con la prueba diagnóstica negativa y, por tanto, sea me-

nos probable que éstos entren en el estudio.

El último criterio que se debe valorar para juzgar la validez del estudio es si se describen los métodos con suficiente detalle como para permitir su reproducción. Esta descripción debe incluir todos los aspectos de la preparación de pacientes, la realización de la prueba y su interpretación.

¿Cuáles son los resultados?

El punto de partida del proceso diagnóstico es habitualmente un paciente

con unas características de gravedad y comorbilidad concretas, que le confieren una probabilidad determinada de tener la entidad a diagnosticar (probabilidad antes de la prueba). El objetivo de la realización de la prueba diagnóstica es, una vez conocido el resultado, modificar esa probabilidad hasta obtener una probabilidad tras la prueba.

Consideremos el escenario diagnóstico más simple, en el que tanto el patrón de referencia como la prueba diagnóstica clasifican a los pacientes en dos grupos en función de la presencia o ausencia de un síntoma, signo o enfermedad. Generalmente, los resultados se expresan a partir de la proporción de aciertos de la prueba diagnóstica entre las poblaciones enferma (sensibilidad) y sana (especificidad). La sensibilidad es la probabilidad de que la prueba sea positiva si la condición de estudio está presente (patrón de referencia positivo), mientras que la especificidad es la probabilidad de que la prueba sea negativa si la condición está ausente. Sin embargo, la sensibilidad o la especificidad no nos facilitan el cálculo de la probabilidad tras la prueba. Para ello resulta más útil el empleo de los cocientes de probabilidades.

El cociente de probabilidades (CP) para un determinado resultado de una

prueba diagnóstica está definido como la probabilidad de dicho resultado en presencia de enfermedad dividida por la probabilidad de dicho resultado en ausencia de enfermedad. Los CP resumen información de la sensibilidad y de la especificidad e indican la capacidad de la prueba para incrementar o disminuir la verosimilitud de un determinado diagnóstico. Utilizando los CP y la probabilidad antes de la prueba podemos calcular las probabilidades tras la prueba de cada paciente individual.

Cuando la prueba que se evalúa se mide en una escala cuantitativa, la sensibilidad y la especificidad varían en función del punto de corte que se utilice para separar los valores normales de los anormales. En estos casos, los resultados pueden representarse gráficamente como una curva ROC, que permite conocer las características de la prueba según diferentes puntos de corte y que puede utilizarse para elegir el más adecuado.

Al igual que en otros tipos de estudios, la valoración de la validez de las pruebas diagnósticas se hace sobre muestras, por lo que los resultados obtenidos son sólo estimaciones puntuales, sujetas a variabilidad aleatoria y, por tanto, deben darse con sus intervalos de confianza.

¿Son aplicables en tu medio?

A la hora de valorar la aplicabilidad de la prueba diagnóstica en nuestros pacientes tenemos que considerar si la prueba es suficientemente reproducible, independientemente del ámbito y las condiciones de aplicación. Por ello, resulta importante que los trabajos publicados incluyan información expresa sobre la reproducibilidad de la prueba y sus criterios de interpretación.

Otro aspecto que se debe tener en cuenta es si resulta razonable asumir que las características operativas de la prueba diagnóstica, estimadas en el estudio, van a ser similares en nuestros pacientes. Si el ámbito en el que se ha evaluado la prueba es similar al nuestro, nuestros pacientes cumplen los criterios de inclusión del estudio y no violan los de exclusión, parece sensato aceptar la aplicabilidad de los resultados. Si el espectro de pacientes incluido en el estudio es diferente al nuestro, la decisión deberá tomarse con cautela.

Finalmente, la utilidad clínica de la prueba dependerá de la repercusión que tengan sobre nuestra actitud diagnóstica y terapéutica y, sobre todo, de si al margen de que influya en nuestras decisiones, su uso aporta algún beneficio al paciente.

Conclusiones

Hemos revisado los principales criterios que se deben considerar en la valoración de la validez y la aplicabilidad de los estudios de evaluación de intervenciones sanitarias y de pruebas diagnósticas. Son herramientas de gran utilidad para poder resolver problemas clínicos mediante el análisis eficiente de la literatura científica. Herramientas metodológicas que, si adquirimos y ejercitamos, nos ayudarán a valorar cualquier evidencia científica y a integrarla con nuestros conocimientos y experiencia clínica para poder decidir sobre su aplicabilidad e idoneidad en un paciente concreto.

Bibliografía

1. Sackett DI, Richardson WS, Rosenberg W, Haynes RB. Medicina basada en la evidencia. Cómo ejercer y enseñar la MBE. Madrid: Churchill Livingstone; 1997.
2. Guyatt GH, Sackett DL, Cook DJ. 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;270:2598-601.

3. Guyatt GH, Sackett DL, Cook DJ. 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.

4. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA. 1994;271:389-91.

5. Jaeschke R, Gordon H, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? JAMA. 1994;271:703-7.

