

Oftalmología basada en la evidencia. Protocolo de lectura crítica de artículos sobre tratamiento

JC. Mesa^{1,2}
O. Garcia³
J. Lillo³
F. Mascaró³
J. Arruga¹

¹Doctor en Medicina
²Máster MBE
(Medicina Basada
en Evidencias)
Colaboración Cochrane
³Licenciado en
Medicina
Servicio
de Oftalmología
Hospital Universitari
Bellvitge
l'Hospitalet
de Llobregat
Barcelona

Correspondencia:
Juan Carlos Mesa Gutiérrez
Servicio de Oftalmología
Ciutat Sanitària i
Universitària de Bellvitge
Feixa Llarga, s/n
08907 L'Hospitalet.
Barcelona
E-mail:
juancarlosmesa@lycos.co.uk

Resumen

La medicina basada en la evidencia (MBE) es el uso consciente y juicioso de la mejor evidencia disponible en la toma de decisiones. Se trata de actuar en la clínica utilizando información seleccionada y relevante, y que venga avalada por datos obtenidos a través del método científico más riguroso: la epidemiología y la estadística. La MBE resta fuerza a la intuición, la experiencia clínica no sistematizada y la fisiopatología como elementos suficientes para la toma de decisiones clínicas, y acentúa el valor del examen riguroso de las pruebas científicas suministradas por la investigación clínica. La evaluación crítica de los artículos relacionados con las alternativas terapéuticas constituirá nuestro objetivo. Suministraremos las habilidades básicas para la evaluación y análisis de los artículos científicos mediante el uso de una serie de conocimientos sencillos de estadística, diseño de investigaciones y epidemiología clínica.

Resum

La medicina basada en l'evidència (MBE) és l'ús conscient i judiciós de la millor evidència disponible en la presa de decisions. Es tracta d'actuar en la clínica utilitzant informació seleccionada i rellevant, i que vingui avalada amb dades obtingudes mitjançant el mètode científic més rigorós: l'epidemiologia i l'estadística. La MBE resta força a la intuïció, l'experiència clínica no sistematitzada i la fisiopatologia com a elements suficients per a la presa de decisions clíniques, i accentua el valor de l'examen rigorós de les proves científiques subministrades per la investigació clínica. L'avaluació crítica dels articles relacionats amb les alternatives terapèutiques constituirà el nostre objectiu. Subministrarem les habilitats bàsiques per a l'avaluació i anàlisi dels articles científics mitjançant l'ús d'una sèrie de coneixements senzills d'estadística, disseny d'investigacions i epidemiologia clínica.

Summary

Evidence-based-medicine (EBM) is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. It deals with selected and relevant information, supported by data obtained through the most rigorous scientific method: Epidemiology and Statistics. EBM does not rely on intuition, non-systematic clinical experience or Physiopathology as the only elements when taking clinical decisions and stress the value of a thorough examination of scientific evidence given by clinical investigation. Critical evaluation of papers related to diagnostic tests and treatment modalities will be our goals. We will provide with the basic skills for evaluation and analysis of papers by means of amenable topics on Statistics, clinical assays and Epidemiology.

Definición y antecedentes

La medicina basada en la evidencia (MBE) es el uso consciente, explícito y juicioso de la mejor evidencia disponible en la toma de decisiones sobre los cuidados de pacientes individuales¹. Este término fue acuñado en los años 80 por un grupo de internistas y epidemiólogos clínicos canadienses de la Universidad de McMaster, que más tarde formarían

el Evidence-Based Medicine Working Group y su difusión en la práctica clínica se produjo a partir de 1992 con la serie de artículos publicados en la revista JAMA². Su sistemática de trabajo ha ido progresivamente arraigando en la comunidad médica, y el resultado se ha consolidado como un nuevo paradigma o "estilo del saber médico"³ acerca de los conocimientos necesarios para orientar la práctica clínica.

La práctica médica tradicional, heredera de los paradigmas clásicos, pero vigente todavía hoy en muchos ámbitos, se puede caracterizar por la creencia en que:

- Las observaciones derivadas de la experiencia clínica personal son una forma válida de generar, validar y transmitir los conocimientos acerca del pronóstico de las enfermedades, el rendimiento clínico de las pruebas diagnósticas y la eficacia o eficiencia de los tratamientos.
- Un buen conocimiento de la teoría fisiopatológica subyacente, una combinación de habilidad en el razonamiento y la especulación lógica, y una buena dosis de sentido común permiten interpretar adecuadamente los signos de la enfermedad y elegir el tratamiento más adecuado.
- Los conocimientos se actualizan adecuadamente mediante libros de texto y revistas con los tradicionales "artículos de revisión", en los que los expertos de más experiencia y prestigio nos seducen con sus opiniones juiciosas acerca de las soluciones a los problemas clínicos. Es por ello que se concede una enorme credibilidad al argumento de autoridad, y los apartados "introducción" y "discusión" de los trabajos de investigación original centran los debates y marcan las pautas de actuación para una "buena praxis".

Sin embargo, la práctica clínica diaria nos sigue demostrando el error de las premisas anteriores y su insuficiencia para guiar nuestras decisiones cotidianas. Debemos aceptar que desconocemos el impacto real de muchas de las medidas que tomamos en nuestro quehacer cotidiano, pues su eficiencia clínica nos parece intrínseca a su coherencia con la base fisiopatológica teórica del problema. Un ejemplo claro de ello nos lo ha proporcionado el uso de la seroalbúmina en el tratamiento de pacientes hipoproteinémicos. Parece evidente que en pacientes gravemente hipoalbuminémicos tenga sentido fisiopatológico administrar albúmina intravenosa para restaurar la cifra de esta proteína plasmática y así mejorar el pronóstico del paciente evitando la formación de edemas. Pero una revisión sistemática de la literatura ha demostrado que en pacientes hipoproteinémicos (y también en quemados), el uso de albúmina para restaurar sus niveles plasmáticos no sólo no mejora el pronóstico, sino que produce mayor mortalidad⁴. En concreto, hay que administrar albúmina a 16 pacientes hipoproteinémicos para producir una muerte. Si esto no es coherente con la hipotética base fisiopatológica, el método científico que hemos desarrollado desde el Renacimiento nos ha enseñado que probablemente lo erróneo sea la teoría que nuestra mente ha construido para explicar

el problema. Con un experimento sencillo fue como, en el siglo XVI, Harvey demostró que las teorías de Galeno sobre la circulación -inmutablemente vigentes desde el imperio romano- eran falsas, inaugurando la "nueva práctica" de una medicina científica⁵.

La historia del uso del suero frío en la hemorragia digestiva nos ilustra de lo endeble que resulta basar nuestra práctica clínica en el "argumento de autoridad". Esta medida terapéutica alcanzó una gran difusión una vez que fue planteada por Wangensteen a finales de los 50, a partir de sus trabajos experimentales con perros. A pesar de los deficientes resultados conseguidos en la clínica, de trabajos posteriores en los que se observaba un efecto perjudicial del frío sobre la mucosa del estómago y sobre la coagulación, y de que los resultados eran similares si se utilizaba agua corriente, los tratados clásicos de cirugía de los años 80 seguían planteando esta opción en el tratamiento empírico de la hemorragia digestiva alta⁶. También el sistema tradicional de reciclaje de conocimientos, el famoso sistema de aprendizaje mediante sesiones de "Formación Médica Continuada", ha quedado obsoleto, puesto que los libros de texto tradicionales son incapaces de recoger la nueva información científica que se produce en tiempo real. Así, después de 43 ensayos clínicos aleatorizados (en los que participaron más de 21.000 pacientes) en que se demostraba la eficacia de la trombólisis temprana sobre la mortalidad del infarto de miocardio, ningún libro de texto médico establecía aún esta indicación como rutinaria. Pero aún hay más: al igual que en el caso del suero frío, en 1990 y después de 15 ensayos clínicos aleatorizados y tres metaanálisis, se seguía recomendando en los textos especializados la administración profiláctica de lidocaína para prevenir el re-infarto, un medida completamente ineficaz⁷.

Por otra parte, en las últimas décadas hemos asistido al fenómeno de la explosión informativa, un crecimiento exponencial de la literatura médica que hace materialmente imposible mantenerse al día si pretendemos hacerlo utilizando este recurso de una manera acrítica. En 1948 había cerca de 4.700 revistas científicas; en 1994 se publicaban unos 2 millones de artículos en 20.000 revistas médicas⁸. Aunque quizá en nuestra especialidad sea más sencillo dado (por el momento) el menor volumen de información a consumir, se estimó que para mantenerse al día, un generalista debería leer 19 artículos al día durante los 365 días del año⁹. Si combinamos este hecho con la mejora actual en el acceso a la información gracias internet, el resultado es una avalancha de información que obliga al médico que no quiera ir perdiendo competencia profesional con el paso del tiempo, a dominar habilidades y técnicas sistemá-

ticas que le doten de sentido crítico y le permitan identificar la información verdaderamente relevante para su práctica “a pie de paciente”. Necesitamos el imán que nos permita buscar la aguja en el pajar, puesto que con el actual ritmo de producción de ensayos clínicos y otras investigaciones rigurosas, la cuestión ha dejado de ser si nuestras actuaciones en la práctica tienen buena base científica, sino cuánta de la evidencia actualmente disponible se aplica en la práctica diaria¹⁰.

Frente al modelo tradicional, la MBE se caracteriza por la creencia en que:

- La información derivada de la experiencia clínica y la intuición puede llevar a conclusiones erróneas si no se basa sólidamente en observaciones sistemáticas.
- El estudio y conocimiento de los mecanismos teóricos básicos de la enfermedad es necesario pero insuficiente para guiar la práctica clínica.
- El profesional necesita conocer ciertas reglas para evaluar rigurosamente la metodología con la que se han obtenido las pruebas científicas en las que se sustentan sus decisiones.

La MBE resta fuerza a la intuición, la experiencia clínica no sistematizada y la fisiopatología como elementos suficientes para la toma de decisiones clínicas, y acentúa el valor del examen riguroso de las pruebas científicas suministradas por la investigación clínica. Para ello incorpora al arsenal de saberes y habilidades básicas para el desempeño de la profesión médica, la destreza en el uso de una serie de conocimientos sencillos de estadística, diseño de investigaciones y epidemiología clínica.

Los médicos deben, pues, adquirir la responsabilidad de evaluar de forma crítica e independiente la credibilidad tanto de las evidencias como de las opiniones ofertadas. Lo importante no es el mensaje, sino el método con el que se ha llegado a los datos. Para los científicos la validez del conocimiento teórico radica en si resiste la dura prueba que supone el experimento. No importa lo maravillosa, ingeniosa o coherente que nos parezca la conjetura que hemos supuesto, ni cómo se llama –o qué cargo ocupa– el que la formuló: si un experimento contradice a la teoría, la teoría es falsa¹¹. Son ahora los apartados “material y métodos” y “resultados” de los artículos de investigación los que se convierten en las piezas claves de los trabajos médicos, pues son las secciones que deben evaluarse detalladamente para valorar la validez de los datos que aporta. Así que para este nuevo estilo científico de hacer medicina, la autoridad establecida –los “expertos”–, tienen mucho menor

peso. Por eso Sackett¹², convertido a su vez en experto muy a su pesar, aboga por la desaparición de esta figura para facilitar el avance de la ciencia: primero por la tendencia existente en el resto de la comunidad médica a no contradecirlos, ya sea por deferencia, miedo o respeto; segundo, porque los editores de las revistas se enfrentan a la tentación de aceptar o rechazar nuevas ideas y evidencias en función de su coincidencia o no con la opinión “experta”.

Esto no debe interpretarse como un rechazo a lo que uno puede aprender de sus maestros o colegas. Únicamente significa que, si se busca la mejor atención para nuestros enfermos, una “buena praxis” de la medicina moderna debe necesariamente partir de un conocimiento riguroso de las pruebas científicas que sustentan cada una de sus prácticas clínicas.

El “arte” de la medicina consistirá en saber combinar ese conocimiento con la valoración sopesada de si con ello se cumplen las expectativas de los pacientes, que son las que finalmente dan sentido a nuestras actuaciones.

La práctica de la MBE

Así pues, se trata de actuar en la clínica utilizando información seleccionada y relevante, y que venga avalada por datos obtenidos a través del método científico más riguroso: la epidemiología y la estadística. Ello no significa que haya que ser un entendido en epidemiología o en estadística para aplicar los principios de la MBE: es factible adquirir unas habilidades básicas que nos posibiliten tener juicio crítico para obtener la mejor evidencia científica del tema que nos interese (la mejor evidencia disponible, puesto que no en todos los campos existe evidencia de primera clase, sobre todo si la producción científica en el área de interés está dominada por las series de casos¹³).

La evidencia conseguida debe integrarse con la experiencia clínica individual y las expectativas, preferencias y deseos del paciente. Su práctica empieza y termina con el paciente.

La actividad clínica diaria genera cuestiones acerca de los efectos de la terapia, la utilidad de una prueba diagnóstica, el pronóstico de una enfermedad o la etiología de una determinada patología, lo que nos lleva a plantearnos una pregunta clínica (paso 1). A continuación realizaríamos una búsqueda bibliográfica (paso 2), evaluaríamos su validez y aplicabilidad (paso 3), y de nuevo volveríamos al paciente, integrando la evidencia con la experiencia clínica y sus preferencias

(paso 4), evaluando el rendimiento de nuestra aplicación y cerrando el círculo de la MBE (Tabla 1).

La pregunta clínica

Se pueden formular dos tipos de preguntas: las denominadas básicas constan de dos componentes y se plantean en relación al conocimiento general de un tema. Si me enfrento a un niño de 5 años con diagnóstico de retinoblastoma y habitualmente no trabajo con pacientes oncológicos, lo más probable es que me plantee una pregunta de dos componentes, (p. ej., tratamiento del retinoblastoma en un escolar), y para responderla podría acudir a un libro de texto clásico sobre oftalmología pediátrica. Sin embargo, la pregunta necesaria para la práctica de la MBE consta como mínimo de 3 elementos y se resume en los acrónimos PIO o PICO (Paciente, Intervención, Comparación, Resultado-Outcome en inglés), y no puede responderse recurriendo a un libro de texto, sino a artículos o revisiones sistemáticas. Este tipo de preguntas son las que se nos plantean en la práctica diaria en temas que sí conocemos y con los que trabajamos habitualmente. La C del acrónimo, la "comparación" no siempre es necesaria. Una pregunta clínica bien formulada va a facilitar enormemente

la búsqueda de la evidencia al permitirnos traducir fácilmente nuestros términos a palabras clave (descriptores) (Tabla 2).

Una vez formulada la pregunta, debemos considerar qué aspecto clínico trata la pregunta: tratamiento o prevención, pronóstico, causalidad-etilogía, elección de una prueba diagnóstica, riesgo-beneficio, calidad de vida, etc., y en función de esto sabremos qué tipo de estudio es el que mejor responde a la pregunta. En la Tabla 3 se presentan los estudios adecuados en función de la pregunta a contestar. No debe confundirse la pirámide de la evidencia aportada por los distintos tipos de estudios (Figura 1) con la adecuación para responder preguntas clínicas. Así, es cierto que un metaanálisis de ensayos clínicos aleatorizados (ECA) aporta más evidencia que un ensayo clínico aislado, y éste a su vez es superior a un estudio de cohortes, y éste es superior a un caso-control, etc., pero un ECA no es el mejor estudio para valorar una prueba diagnóstica, donde no es necesario aleatorizar a los pacientes, sino realizar comparaciones independientes de la prueba a estudiar con la prueba de referencia (patrón oro); tampoco se presenta el ECA como el estudio a buscar en caso de factores de riesgo, puesto que es absurdo pensar en un ensayo clínico en el que los pacientes sean sometidos al azar a un posible factor de riesgo.

El paciente	Se plantea una cuestión acerca del pronóstico, el tratamiento, una prueba diagnóstica.
La pregunta	Definir un problema clínico (pregunta) y la información necesaria para resolverlo.
La búsqueda	A partir de distintas bases de datos se obtiene la bibliografía seleccionada.
La evaluación	Se valora la validez metodológica de los trabajos y se evalúa su validez y aplicabilidad.
El paciente	La evidencia obtenida debe aplicarse en combinación con nuestra experiencia y las preferencias del paciente.
Autoevaluación	Se debe evaluar el resultado obtenido tras la aplicación del conocimiento obtenido.

Tabla 1.
El proceso de la MBE

La pregunta clínica			
P	Paciente	¿Cómo describe al paciente que está tratando? Sea preciso	Hombre 70 años con cuadro funduscópico compatible con DMAE húmeda, con MNV yuxtafoveal predominantemente clásica.
I	Intervención	¿Cuál es la intervención principal que está considerando? Maniobra terapéutica, realización de una prueba diagnóstica, pronóstico de una enfermedad.	Terapia: Inyección intravítrea de fármacos anti-angiogénicos
C	Comparación (si procede)	¿Cuál es la principal alternativa con la que comparar la intervención?	Terapia fotodinámica con Verteporfin
O	Outcome (resultado)	¿Qué espera conseguir, mejorar o en qué puede afectar la medida tomada?	Resultado 1: Mejora de agudeza visual. Resultado 2: Mantenimiento de agudeza visual

Tabla 2.
Componentes de la pregunta clínica

La búsqueda de información

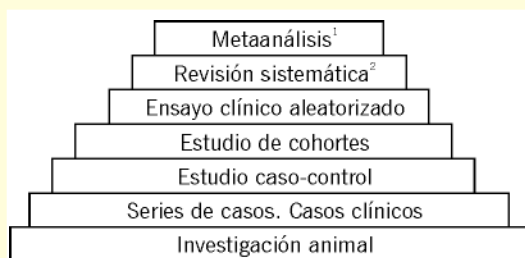
Las fuentes para responder a nuestra pregunta clínica son varias. Podemos recurrir a libros de texto tradicionales pero, como hemos visto, la información que contienen queda obsoleta con rapidez y no son adecuados para responder a preguntas de 3 componentes.

Se puede recurrir a bases de datos con filtro de calidad, como *Embase*, base de datos del repertorio Excerpta Medica; o la popular *Medline* (www.ncbi.nlm.nih.gov), base de datos del repertorio Index Medicus, producido por la National Library of Medicine y de libre acceso, a diferencia de *Embase*, gracias a la administración Clinton. Aunque en cualquier biblioteca podemos encontrar asesoramiento acerca de cómo realizar una búsqueda en *Medline*, en caso de no estar familiarizado con la mecánica se puede obtener un manual de trabajo en castellano a través de la página web www.fisterra.com, dedicada a atención primaria (http://www.fisterra.com/recursos_web/

no_explor/pubmed.asp). En cualquier caso, la propia base de datos posee un buscador de preguntas clínicas (Clinical Queries) donde se pueden introducir los términos de búsqueda para terapia, diagnóstico, pronóstico o etiología, facilitando enormemente la realización de este tipo de búsquedas.

A partir de esta base de datos obtendríamos los artículos de interés, pero éstos deberán ser evaluados para valorar la evidencia que aportan. Una tercera alternativa es la búsqueda en revistas secundarias, que realizan el proceso evaluador por nosotros y nos ofrecen información ya revisada y catalogada desde el punto de vista de la evidencia a partir de artículos metodológicamente sólidos. Hoy en día existen en la red diferentes fuentes que proporcionan información de este tipo. En castellano se puede acceder a ellas a través de páginas como www.fisterra.com, o <http://infodoctor.org>, que contiene la página de Rafael Bravo, con múltiples recursos (<http://infodoctor.org/rafabravo>). La página web de la Universidad de Washington ofrece vínculos a distintas fuentes de MBE, incluyendo la colaboración Cochrane (<http://healthlinks.washington.edu/ebp/ebpresources.html>).

Figura 1.
La pirámide de la evidencia



¹Estudio que mediante técnicas estadísticas combina los resultados de varios estudios del mismo tipo (generalmente ensayos clínicos o estudios de cohorte), obteniendo así un estudio con mayor muestra de población y, por tanto, mayor potencia estadística.

²Estudios centrados en una pregunta clínica concreta y elaborados a partir de búsquedas extensivas de la literatura, seleccionando sólo aquellos artículos metodológicamente correctos, cuyo contenido se evalúa y resume, presentando una respuesta concreta a la pregunta clínica en cuestión.

Esta forma de utilización de la literatura médica en base a problemas clínicos constituye una medida más eficaz que la lectura tradicional para mantenerse al día y actualizar los conocimientos en la especialidad¹⁴.

Evaluación crítica de artículos relacionados con el tratamiento

Un artículo sobre tratamiento debe responder a 3 aspectos: validez de los resultados, importancia de los mismos y aplicabilidad a pacientes individuales. Analizamos y proporcionamos las herramientas para responder a estas preguntas: la evaluación de la validez interna, que garantiza que los resultados del estudio no están sesgados; el efecto del tratamiento y utilidad para nuestro trabajo diario.

Tabla 3.
Tipo de estudio en función del aspecto clínico a responder

Aspecto clínico	Estudio
Tratamiento-Prevención	Ensayo clínico aleatorizado (ECA) o revisión sistemática (metaanálisis) de ECA.
Pronóstico	Cohortes incipientes ("de inicio") con análisis de supervivencia.
Factores de riesgo-Etiología	Cohortes o caso-control con análisis multivariante.
Diagnóstico	Transversal: comparación independiente con patrón oro.
Motivación	Cualitativo

La literatura médica ha de ser evaluada para valorar su nivel de evidencia, labor que realizan por nosotros las denominadas revistas secundarias, como el ACP Journal Club, la colaboración Cochrane o, más frecuentemente, nosotros mismos con la ayuda de las llamadas guías de lectura crítica. Esta labor de “evaluación”, aunque pueda parecer farragosa en principio, nos permite, por un lado, conocer de forma rápida si merece la pena dedicar tiempo a un artículo y, por otro, limitar el número de artículos necesarios para mantenernos al día. Las guías de lectura crítica no son más que listas de verificación que deben cumplimentarse con los datos que obtenemos del estudio y que nos ayudarán a decidir sobre su idoneidad para responder a una pregunta clínica. En la Tabla 4 se presenta la guía de lectura crítica que vamos a utilizar para los artículos sobre tratamiento.

Pueden obtenerse guías similares en las páginas web de la red CASP España, (Critical Appraisal Skills Programme)¹⁵ o del centro para la medicina basada en la evidencia de la Universidad de Oxford¹⁶.

¿En qué situaciones nos vamos a plantear la realización de una lectura crítica? Generalmente cuando queramos responder a una pregunta clínica determinada. Un ejemplo: en el último congreso de la Sociedad Española de Oftalmología, Vd. presencia una comunicación sobre la conveniencia de tratamiento de la hipertensión ocular mediante fármacos

antihipertensivos, tras la que se establece cierta discusión. A la vuelta, y puesto que la discusión se estableció en base a una serie de casos, decide buscar si existe una evidencia más sólida. Vd. sabe que la máxima evidencia sobre la eficacia de un tratamiento se consigue mediante un ensayo clínico aleatorizado, y realiza una búsqueda bibliográfica a través de MEDLINE utilizando como palabras clave “ocular hypertension”, “medical treatment” limitando la búsqueda utilizando como tipo de artículo la opción “randomized controlled trial”. El resultado son 18 referencias bibliográficas entre las que se incluyen ensayos con distintos fármacos, comparaciones entre ellos y también un ensayo multicéntrico publicado en Archives of Ophthalmology¹⁷, revista disponible en la biblioteca de su centro hospitalario.

Lectura crítica ¿qué debemos pedirle a un artículo sobre tratamiento?

Las tres preguntas genéricas a las que debemos responder son:

- ¿Son válidos los resultados del estudio?
- ¿Cuál es la importancia clínica de los resultados?
- ¿Son aplicables los resultados a mis pacientes?

¿Son válidos los resultados del estudio?

Criterios primarios

- ¿Existe una pregunta clínica claramente definida?
- ¿La asignación de los pacientes fue aleatoria?
- ¿Se mantuvo oculta la secuencia de aleatorización?
- ¿El seguimiento de los pacientes fue completo?
- ¿Los pacientes se analizaron en el grupo al que fueron asignados?

Criterios secundarios

- ¿Se ha mantenido un diseño ciego respecto a los pacientes, los clínicos y el personal de estudio?
- ¿Los grupos eran similares al inicio del ensayo?
- Aparte de la intervención, ¿se han tratado a los dos grupos de la misma forma?

¿Cuáles han sido los resultados?

- ¿Cuál ha sido la magnitud del efecto del tratamiento?
- ¿Con qué precisión se ha estimado el efecto?

¿Me resultarán útiles?

- ¿Pueden aplicarse los resultados a mis pacientes?
 - ¿Se han tenido en cuenta todos los resultados de importancia clínica?
 - ¿Compensan los beneficios del tratamiento, los posibles efectos adversos y los costes?
-

Tabla 4.
**Guía de lectura crítica
para artículos sobre
tratamiento**

En este artículo nos centraremos en responder la primera cuestión, es decir, en evaluar su validez, entendiendo como tal su validez interna, que garantiza que sus resultados no estén sesgados.

Análisis de la validez

Identificación de la pregunta clínica

La identificación de los elementos de la pregunta clínica nos permitirá, además de evaluar su validez interna (ausencia de sesgo), conocer su aplicabilidad, es decir, la generalización de sus resultados a nuestros pacientes, en el caso de que la muestra utilizada sea representativa. En la Tabla 5 se muestran los criterios de inclusión en el estudio elegido, las intervenciones a comparar y los resultados registrados, siendo el principal el desarrollo de un defecto en el campo visual o la afectación de la papila.

La importancia de la aleatorización

La segunda cuestión a responder, y posiblemente la más importante, es si el ensayo ha sido aleatorizado. Varias son las razones por las que es necesario que en un ensayo clínico exista una distribución aleatoria. La primera de ellas es que el muestreo aleatorio es imprescindible para que se cumplan los requisitos matemáticos de aplicación de los tests estadísticos. Toda la inferencia estadística y el cálculo de probabilidades se basan en el estudio de muestras extraídas al azar. La aleatorización es el único momento del diseño experimental en el que se introducen explícitamente las leyes del azar: ellas rigen la distribución de frecuencias teóricas que vamos a comparar con las frecuencias observadas. Así pues, sólo la aleatorización da sentido real al

uso de los tests estadísticos (buscar la «p» o el intervalo de confianza (IC) al 95%), que se utilizan para medir el poder del azar en los resultados que se están encontrando.

La segunda es que la aleatorización de muestras grandes es el único método conocido de evitar el sesgo de confusión. Si en un experimento partimos inicialmente de dos muestras idénticas, y sobre una de ellas (pero no sobre la otra) realizamos una intervención, si al final aparecen diferencias entre ambas muestras podemos afirmar que la causa de las diferencias es la intervención. Si por el contrario, en un experimento partimos inicialmente de muestras que no son iguales, después de realizar la intervención sólo en una de ellas, si al final aparecen diferencias entre ellas no podremos achacar la causalidad de esas diferencias a la intervención. Así que la igualdad de las muestras es un requisito imprescindible para atribuir la causalidad, para concluir que la causa de las diferencias en el resultado final es la única variable distinta entre los grupos: la intervención que se está estudiando.

En base al teorema denominado la “Ley de grandes números” (LGN)¹⁸, la aleatorización de muestras grandes tiende a producir grupos uniformes en todas las variables (incluidas las desconocidas), previamente a que se aplique la intervención en estudio. A consecuencia de esta ley, el control de las variables de confusión será proporcional al tamaño muestral: el teorema funciona si el muestreo es aleatorio cuando n (tamaño muestral) es ∞ . Como en nuestros ensayos clínicos nunca vamos a disponer de muestras infinitas, es imprescindible demostrar a los lectores de nuestros estudios que el efecto de la LGN para homogeneizar las muestras y controlar el efecto de los confusores principales se ha cumplido para el tamaño muestral que hemos elegido. Para ello sirven las llamadas “Tabla 1” de los trabajos: por consenso¹⁹ se ha establecido que la primera tabla de un ensayo clínico debe mostrar la frecuencia de aparición de las principales variables demográficas y/o de confusión en ambas muestras antes de que se aplique la intervención. En esta Tabla no encontraremos ninguna prueba estadística para comparar ambos grupos, ninguna p que nos diga que no existen diferencias estadísticamente significativas. La p no mide la homogeneidad, además de que una pequeña diferencia entre dos grupos alcanzaría significación estadística si el tamaño muestral fuera lo suficientemente grande. Es el lector el que, a la vista de los resultados y aplicando su juicio, evaluará si los grupos son comparables. En la Tabla 6 se muestra la “Tabla 1” del ensayo que hemos elegido. Nótese el poder del azar para generar

Tabla 5.
Respuesta a la pregunta
clínica del OHTS

Participantes	1636 pacientes entre 40-80 años, con PIO 24-32 mmHg en un ojo y 21-32 mmHg en el contralateral; con ángulo abierto, 2 campos visuales previos normales y con fotografía estereoscópica de la papila normal.
Intervenciones	Tratamiento tópico antihipertensivo hasta disminuir la PIO un 20% y <24 mmHg.
Resultados	Reducción de PIO. Incidencia acumulada de glaucoma primario de ángulo abierto. Seguimiento 78 meses.

dos muestras muy similares, aunque no idénticas, como en el caso de los pacientes con antecedentes familiares de glaucoma o la asociación con miopía e hipertensión arterial, más numerosos en el grupo tratado con antihipertensivos tópicos. Usted, lector, tendrá que evaluar en base a sus conocimientos y experiencia si esa diferencia es significativa o puede influir en el resultado.

¿Cómo comprobar que se ha realizado una correcta aleatorización?

El término *muestra aleatoria* tiene una estricta acepción matemática que difiere del que se usa popularmente. En teoría, para que una muestra sea aleatoria se requiere que cada uno de los miembros de la población que se muestrea tenga la misma

Característica	Medicación (n=817)	Observación (n=819)	Total (n=1636)
Sexo, n %			
H	359(43.9)	346(42.2)	705(43.1)
M	458(56.1)	473(57.8)	931(56.9)
Edad, n %			
40<=50	291(35.6)	287(35)	578(35.3)
>50 <=60	270(33.0)	259(31.6)	529(32.3)
60-70	202(24.7)	210(25.6)	412(25.6)
70-80	54(6.6)	63(7.7)	117(7.2)
Raza, n %			
Americano nativo	1(0.1)	3(0.4)	4(0.2)
Asiático	4(0.5)	10(1.2)	14(0.9)
Afroamericano	203(25)	205(25)	408(25)
Hispano	24(2.9)	35(4.3)	59(3.6)
Blanco	577(70.6)	560(68.4)	1137(69.5)
Otros	8(1)	6(0.7)	14(0.9)
PIO; Media (DE), mmHg	24.9(2.6)	24.9(2.7)	24.9(2.7)
Relación excavación/disco horizontal, DM	0.39(0.19)	0.36(0.18)	0.36(0.18)
Relación excavación/disco vertical, DM	0.39(0.20)	0.39(0.19)	0.39(0.19)
CV, DM, media (DE) (dB)	+0.27(1.07)	+0.21(1.03)	+0.24(1.05)
CV, Patrón desviación (Db), media (DE), dB	1.92(0.21)	1.90(0.21)	1.91(0.21)
CV, patrón desviación corregida, media (DE), dB	1.12(0.34)	1.12(0.36)	1.12(0.35)
Espesor corneal central, media (DE), micras	570.5(38.9)	274.5(37.7)	572.5(38.4)
Uso previo de medicación hipotensora ocular, %	35.0	39.3	37.2
Historia de glaucoma en familiar de primer grado	34.0	35.6	34.8
Miopía EE>1D, %	34.4	33.7	34.1
β -bloqueantes orales,%	5.4	4.6	5.0
Antagonistas canales de calcio orales	12.8	14.0	13.4
Antecedentes de migraña,%	10.4	11.7	11.1
Antecedentes de diabetes,%	11.5	12.1	11.8
Antecedentes hipertensión,%	37.5	38.1	37.8
Antecedentes hipotensión,%	4.8	4.0	4.4
Antecedentes enfermedad cardiovascular,%	5.8	6.5	6.1
Antecedentes de infarto%	0.9	1.6	1.2

Tabla 6.
Respuesta a la pregunta clínica del OHTS

probabilidad de ser seleccionado para incluirse en la muestra: la probabilidad de asignación a los diferentes grupos es *fija e igual* para todos y cada uno de los individuos que participan. Este requerimiento no se cumple si la muestra se elige de manera fortuita (esto es, sin un plan específico) o de manera sistemática (números pares o impares de la historia clínica o días alternos de acudir a consulta) o conveniente (todos los que han devuelto un cuestionario contestado). Para obtener una muestra verdaderamente aleatoria de una población de pacientes se requieren maniobras especiales, como tirar una moneda insesgada, utilizar una tabla de números aleatorios o un generador computerizado de números aleatorios²⁰.

Como hemos visto, la aleatorización es la parte más sensible de todo el diseño. Por ello, para dar validez a los resultados del ensayo, deberemos aprender algunos “trucos” para detectar si la aleatorización ha sido realizada correctamente. El primero, que ya hemos citado, es comprobar que el azar ha generado muestras idénticas cotejando los datos de la llamada “Tabla 1”. Pero quizá lo más curioso es saber que el muestreo aleatorio simple tiende a producir muestras homogéneas, *pero de tamaños muestrales diferentes*. La mayoría de la gente piensa que, por ejemplo, cuando distribuimos al azar en dos grupos (probabilidad de pertenencia a cada grupo = 0,5) un muestra total de 40 pacientes, cada una de las dos muestras resultantes deben poseer 20 pacientes. Nada más lejos de la realidad, pues la teoría de probabilidades nos permite calcular, utilizando la distribución binomial (Figura 2), que la probabilidad de obtener ese resultado es del 12,54%.

Es sorprendente encontrar en la literatura el enorme número de ensayos clínicos pequeños (menos de 100 pacientes) en los que el tamaño de los grupos que se comparan es igual cuando la muestra es un número par (30/30, probabilidad de presentación = 10,3%; 40/40, probabilidad = 8,9%) o hay un desequilibrio de uno cuando la muestra total es impar (17/16, probabilidad = 13,6%; 25/26, probabilidad 11%), cuando, sin embargo, dicen en los métodos que la distribución ha sido aleatoria. El hecho de que contradigan la teoría de probabilidades hace que estos

trabajos sean sospechosos de que posiblemente no se ha realizado una aleatorización real, aunque sus autores así lo afirmen.

Es tan conocido entre los matemáticos que el muestreo aleatorio tiende a producir grupos de tamaños muestrales desiguales, que los estadísticos que trabajan con muestras pequeñas suelen utilizar muestreos especiales como el muestreo por bloques permutados balanceados. Por ello, un buen consejo para detectar que la aleatorización se ha realizado efectivamente es que, si se ha utilizado muestreo aleatorio simple, los grupos deben tener tamaño diferente. Si los grupos son iguales, se debe especificar en los métodos el sistema específico de aleatorización que se ha utilizado (p. ej., bloques balanceados).

En cualquier caso, ya hay pruebas empíricas⁹ de que el factor que más sesga los resultados, sea cuál sea el método de aleatorización utilizado, es que no se respete la *ocultación de la secuencia de aleatorización* (OSA). La OSA (en inglés *allocation concealment*), un concepto diferente del enmascaramiento, consiste en que una vez ha entrado al estudio el primer paciente, debe ser imposible que el personal que administra la intervención adivine el grupo al que pertenecerá el próximo paciente que entre al estudio. Para que la aleatorización sea correcta debe asegurarse la OSA, y para ello los investigadores suelen utilizar varios métodos. El mejor es la aleatorización centralizada (vía telefónica o vía internet) en un centro coordinador diferente al resto de centros que aportan pacientes al ensayo, pero otros métodos válidos de OSA son la utilización de sobres opacos lacrados que contienen la intervención (o su etiqueta, si hay enmascaramiento, etc.) o, si se utiliza para la aleatorización el método de *bloques balanceados*, cambiar el tamaño de los bloques.

Una cuestión secundaria es el diseño ciego respecto al tratamiento. En un estudio ideal, tanto el paciente como el clínico, como el analista de los datos, desconocerían a qué grupo pertenece el paciente, pero esto no va a ser siempre posible. En el caso de terapias quirúrgicas frente a médicas es un buen ejemplo; o tratamiento médico frente a observación, como ocurre en el artículo que hemos seleccionado. En estos casos, al menos los autores habrán intentado que el analista de los resultados desconozca a qué grupo pertenecen los pacientes.

Figura 2.
Ley binomial de la probabilidad

$$P(k) = [n!/k!(n-k)!] \pi^k (1-\pi)^{n-k}$$

P(k): probabilidad de obtener k sujetos.
n: tamaño muestral total.

Seguimiento completo y análisis por intención de tratar

En segundo lugar, el seguimiento que se haya hecho de los pacientes deberá ser completo, es decir, que

todo paciente reclutado para el estudio debe ser tenido en cuenta a su finalización, aunque en cualquier estudio van a existir pérdidas en el seguimiento. Si las pérdidas de seguimiento son excesivas, los resultados del estudio pueden ser puestos en duda, puesto que es frecuente que el pronóstico de los pacientes para los que no se dispone de seguimiento sea diferente al de los demás. ¿Cómo saber si las pérdidas de un estudio invalidan el resultado? Mediante el *análisis de sensibilidad (worst case analysis)*. Consiste en suponer que todos los pacientes perdidos en el grupo control han evolucionado bien, y todos los perdidos en el grupo tratado han evolucionado mal, y volver a calcular los resultados. Si no se modifican los resultados del estudio, las pérdidas son asumibles. En caso contrario, la fuerza de los resultados se debilita en proporción a la probabilidad que existe de que los tratados hayan evolucionado bien y los del grupo control hayan evolucionado mal.

Aplicándolo a nuestro artículo, ambas ramas del estudio presentan 84 y 89 pérdidas de seguimiento a los 78 meses (Figura 3). Para el análisis de sensibilidad supondríamos que todos los perdidos en el grupo de tratamiento han desarrollado glaucoma y que los no tratados en el grupo control no lo han desarrollado. Parece que se modifican los resultados: el riesgo de desarrollar glaucoma en la rama de tratamiento se modifica (0.05 a 0.15%) y en la rama de observación se queda en 0.11%.

Además de las pérdidas de seguimiento, es posible que los pacientes asignados en una u otra rama no tomen su tratamiento, no sea posible aplicarlo, o incluso que haya cambios de una rama a otra. En cualquier caso, el paciente será analizado dentro del grupo al que fue asignado independientemente del tratamiento que haya recibido. Es lo que se conoce como *análisis por intención de tratar*. Aunque en principio pueda parecer que si un paciente no ha recibido el tratamiento asignado debe ser excluido del análisis en esa rama, esto no es así. En muchos estudios, los pacientes que no toman la medicación evolucionan peor, aunque tomen placebo. En el caso de pacientes con patología quirúrgica, puede que algunos no lleguen a ser operados, puede que por gravedad. Si estos pacientes se incluyen en el brazo de control, incluso una operación inútil se mostrará como efectiva, ya que todos los pacientes con peor pronóstico se han asignado al grupo control²¹.

Esta forma de análisis preserva el efecto de la aleatorización, y aproxima los resultados del estudio a la práctica clínica real. De hecho, en nuestro estudio de ejemplo 702 de los 817 pacientes aleatorizados al grupo de tratamiento tópico lo recibieron. En los

115 restantes no fue posible su aplicación, pero se analizaron igualmente en esa rama.

Igualdad de tratamiento al margen de la intervención

Ambos grupos de estudio deben recibir la misma atención. Si se realizara un seguimiento más estricto de una de las ramas podrían detectarse determinados hechos que no «aparecerían» en la otra rama, lo que podría afectar a los resultados.

Las intervenciones distintas al tratamiento en estudio, también denominadas cointervenciones, también pueden ser un problema (p. ej., utilización de beta-bloqueantes sistémicos como consecuencia de una cardiopatía isquémica), sobre todo si los facultativos conocen el tratamiento que están aplicando (no es un estudio a doble ciego) o se autoriza el uso de terapias muy eficaces, y distintas a las estudiadas, a criterio del médico²².

¿Es válido el artículo que hemos elegido?

Como hemos visto, el artículo define de forma concreta pacientes del estudio mediante criterios de inclusión bien definidos, define las alternativas terapéuticas que va a comparar, y el resultado que



Figura 3.
Evolución de los pacientes reclutados en el ensayo clínico

va a medir. De hecho, en la Tabla 2 se presentan todas las variables que se recogen, aunque en el texto se define el evento primario como desarrollo de alteración en el campo visual o en las fotografías estereoscópicas de la papila, quedando los restantes como secundarios.

El estudio se define como aleatorizado. Nos informan de que utilizaron un análisis por intención de tratar y una secuencia de minimización para asegurar el equilibrio en las variables y un proceso de revisión en punto final para distinguir entre los cambios en el campo visual ocasionados por el glaucoma o los originados por otras causas. Se dispuso de central de aleatorización.

El tamaño de ambos grupos es creíble y vemos como el seguimiento fue casi completo, con 84 y 89 pérdidas por rama. El análisis fue por intención de tratar, manteniendo en la rama del tratamiento tópico aquellos casos en los que no se pudo aplicar. Por las características del estudio, fue doble ciego. Nos consta que el analista desconoce la pertenencia a las distintas ramas de tratamiento.

En definitiva, éste sí parece un artículo válido, por lo que seguiríamos con su lectura para responder a las dos siguientes cuestiones: efecto del tratamiento y utilidad para nuestro trabajo diario.

Bibliografía

- Sackett DL, Rosenberg WMC, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.
- Evidence-Based Medicine Working Group. A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
- Lopez Piñero JM, Terrada Ferrandis ML. *Introducción a la medicina*. 1ª ed. Barcelona: Crítica 2000.
- Cochrane Injuries Group Albumin Reviewers. Human albumin administration in critically ill patients: systematic review of randomised controlled trials. *BMJ* 1998;317:235-240. Disponible en internet en: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=9677209>.
- López Piñero JM. *Breve historia de la Medicina*. 1ª ed. Madrid: Alianza Editorial 2000.
- Andrus CH, Ponsk JL. The effects of irrigant temperatura in upper gastrointestinal hemmorrhage: a requiem for iced saline lavage. *Am J Gastroenterol* 1987;82:1062-4.
- Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240-8.
- Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;309:597-9.
- Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine: a new journal to help doctors identify the information they need. *BMJ* 1995;310:1085-6.
- Bodenheimer T. The American health care system -the movement for improved quality in health care. *N Engl J Med* 1999;340:488-92.
- Feynman RP. En busca de nuevas leyes (Messenger Lectures de la Cornell University, 1964). En: Museu de la Ciència de la Fundació «La Caixa», editor. *El carácter de la Ley Física*. Barcelona: Tusquets Editores 2000;165-90.
- Sackett DL. The sins of expertness and a proposal for redemption. *BMJ* 2000;320:1283.
- Weil RJ. The future of surgical research. *PLoS Med* 2004;1:e13. Disponible en internet en www.plosmedicine.org
- Shin JH, Haynes RB, Johnston ME. Effect of problem-based, selfdirected undergraduate education on life-long learning. *Canadian Medical Association Journal* 1993;148:969-76.
- <http://www.redcaspe.org/herramientas/lectura/11ensayo.pdf>
- http://www.cebm.net/worksheet_therapy.asp
- Kass M, Heuer D, Higginbotham E, Johnson C, Keltner J, Miller P, Parrish R, Wilson M, Gordon M. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;120:701-8.
- Rice JA. Limit Theorems: The law of large numbers. En: *Mathematical statistics and data analysis*. 2ª Edición. California. Duxburi Press. 1995.
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 1996;276:637-9.
- Feinstein AR. Principles of medical statistics. 1ª Edición. Florida. Chapman&Hall/CRC. 2002.
- Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? *Int J Epidemiol* 2002;31:1-5.
- Guyatt GH, Sackett DL, Cook DJ. Guías para usuarios de la literatura médica II. Cómo utilizar un artículo sobre tratamiento o prevención. A.¿Son válidos los resultados del estudio? *JAMA* 1993;270:2598-601.