

El método científico: “cómo hacer una lectura crítica de un artículo”

M. Rumayor Zarzuelo, N. García-Arenzana Les, J. Sánchez Díaz, M. Fuentes Ferrer, N. del Prado González, C. Fernández Pérez

Unidad de Apoyo a la Investigación. Servicio de Medicina Preventiva. Hospital Clínico San Carlos. Madrid

Rumayor Zarzuelo M, García-Arenzana Les N, Sánchez Díaz J, Fuentes Ferrer M, del Prado González N, Fernández Pérez C. El método científico: “cómo hacer una lectura crítica de un artículo”. *Cir May Amb* 2010; 15: 66-70.

INTRODUCCIÓN

Durante la última década, dos movimientos han surgido en la medicina, ambos concebidos para mejorar la atención del paciente. El movimiento que busca el humanismo médico para comprender al paciente como persona, centrándose en los valores individuales y el segundo movimiento basado en la evidencia práctica, que aspira a poner a la medicina en manos de expertos que evalúen mejor los datos disponibles en la literatura científica y desarrollen guías clínicas destinadas a estandarizar los procedimientos y terapias (1).

La medicina basada en la evidencia científica (MBE), como corriente o movimiento, tiene su origen en los años 80 en la Escuela de Medicina de la Universidad de McMaster de Ontario (Canadá) por un grupo de clínicos y epidemiólogos (2). Según Sackett, uno de sus fundadores, no es más que un método para practicar y enseñar la medicina clínica; es el empleo consciente, explícito y juicioso de la mejor evidencia actual en la toma de decisiones sobre el cuidado sanitario de los pacientes (3).

La lectura crítica de la literatura científica, como parte del proceso de la MBE, es una herramienta imprescindible para todos los profesionales sanitarios que deben tomar decisiones en su práctica diaria (4).

La realización de la MBE pasa por cuatro etapas:

El problema clínico debe ser el punto principal de partida para la obtención de la respuesta requerida final.

La revisión bibliográfica de la literatura científica con el objetivo de disponer de información actualizada.

El estudio crítico de la evidencia para verificar su validez y utilidad.

La incorporación de los resultados a la práctica clínica (5).

En muchas ocasiones se da por válido y relevante cualquier información que venga avalada por una prestigiosa revista o importante grupo de investigación. Puede tratarse tanto de investigación básica, como epidemiológica o clínica, que en muchas ocasiones puede suponer cambios para la práctica clínica, en el caso de especialidades médicas. Cuando tengamos en nuestras manos un trabajo (normalmente un artículo original, una revisión o un metanálisis) que vayamos a incorporar a nuestro “cuerpo de conocimiento” es acertado plantearse una duda razonable y ser sistemático en la lectura de ese artículo. Debemos utilizar, en la medida de lo posible, una serie de criterios que nos ayuden a discernir si los resultados son válidos e importantes.

Al realizar la lectura crítica de un artículo vamos a identificar cuestiones que tienen que ver con la dimensión de diseño y de los resultados obtenidos, en términos de validez interna y externa, relevancia y utilidad práctica que repasaremos sistemáticamente en la mayoría de los casos según la presentación habitual del trabajo (IMDR: introducción, métodos, resultados y discusión) (3,6). Para poder hacer una correcta lectura es necesario tener unos conocimientos previos sobre metodología de investigación, estadística y la patología o problema que estamos analizando (4).

El objetivo de este artículo es saber evaluar críticamente la validez de un trabajo científico, saber interpretar los resultados y poder determinar su aplicabilidad a cada paciente.

CRITERIOS GENERALES PARA LA LECTURA CRÍTICA DE ARTÍCULOS

¿Qué tenemos que tener en cuenta para interpretar la calidad de la literatura científica?

Unos objetivos del estudio precisos y claros

Un diseño adecuado para la investigación

Unos criterios de la población a estudio adecuados

Un tamaño muestral suficiente para llegar a conclusiones

¿Son tolerables los sesgos? ¿Se podrían haber mejorado de alguna manera?

¿Las conclusiones están acorde con los objetivos y los resultados? (3)

A raíz de los artículos sobre la medicina basada en la evidencia surgió en Inglaterra un programa para la valoración de la lectura científica: Critical Appraisal Skills Programme (CASP) y que en España se llama "Critical Appraisal Skills Programme España" (CASPe). Los detalles de este programa se pueden encontrar en <http://www.redcaspe.org/que-hacemos/herramientas>. El objetivo fue construir una serie de herramientas para poder realizar una correcta lectura de la literatura científica. Cuando leamos críticamente un artículo deberemos intentar responder a las cuestiones planteadas según la presentación clásica del artículo para pasar a decidir sobre las cuestiones de validez, relevancia y utilidad. Respecto a la validez interna del estudio, se tratará de evaluar si los resultados obtenidos se desvían o no de la realidad, es decir si el estudio está libre de sesgos o error sistemático y por tanto tiene validez interna o no. Algunos de los criterios que se presentan se refieren a la relevancia de los resultados, es decir, saber qué resultados se han obtenido y con qué precisión se han estimado, para poder valorar su importancia sanitaria, social o clínica. Finalmente, la dimensión sobre utilidad de los resultados tiene que ver con su aplicabilidad en nuestra experiencia profesional y ello tiene que ver tanto con nuestro contexto como con el contexto en que se ha realizado la investigación cuyos resultados abordamos críticamente, es decir, con su validez externa, entendida como generalización de los resultados.

Las preguntas que nos plantea las Guías CASPe se estructuran en 3 apartados. El primero hace referencia a la validez interna. Las primeras preguntas son de eliminación. Sólo si las respuestas son afirmativas merece la pena continuar con las restantes.

Un segundo apartado de preguntas donde se analizan los resultados y la precisión de los mismos. En el último apartado las preguntas hacen referencia a si los resultados pueden ser generalizados a poblaciones o pacientes.

Para ver si los resultados son válidos y merece la pena seguir adelante con la lectura del artículo nos tenemos que fijar si la metodología aplicada en los diferentes tipos de estudios ha sido aplicada de forma correcta.

Criterios para la lectura crítica de artículos de casos y control

¿Es el estudio de casos y controles una forma adecuada para contestar la pregunta en estas circunstancias?

— Los casos:

¿Definidos de forma precisa?

¿Representativos de la población definida?

¿Se estableció un sistema fiable para la selección de todos los casos?

¿El marco temporal del estudio es relevante en relación a la enfermedad/exposición?

¿Se seleccionó un número suficiente de casos?

¿Tiene potencia estadística?

— Los controles:

¿Representativos de la población definida?

¿Hay algo "especial" que afecta a los controles?

¿Los no respondedores podrían afectar de alguna manera al resto?

¿Han sido seleccionados de forma aleatoria?

¿Se seleccionó un número suficiente de controles?

— La exposición:

¿Se definió claramente y se midió esta de forma precisa?

¿Los autores utilizaron variables objetivas o subjetivas?

¿Las variables reflejan aquello que se supone que tienen que medir?

¿Los métodos de medida fueron similares tanto en los casos como en los controles?

¿La exposición de interés precede al resultado/variable de medida?

— Factores de confusión:

¿Qué factores han tenido en cuenta y cuáles se han omitido? (genéticos, ambientales, socioeconómicos) (7).

Criterios para la lectura crítica de artículos de cohortes

¿Es el estudio de cohortes una forma adecuada para contestar la pregunta en estas circunstancias?

— La cohorte:

¿Reclutamiento adecuado?

¿Representativa de una población definida?

¿Se incluyó a todos los que deberían haberse incluido?

¿La exposición se midió de forma precisa?

¿Las medidas reflejan de forma adecuada aquello que tienen que medir?

¿Sistema fiable de detección de todos los casos?

¿Clasificación de todos los sujetos en el grupo exposición utilizando el mismo tratamiento?

¿Los métodos de medida fueron similares en los diferentes grupos?

¿El seguimiento de los sujetos fue adecuado?

¿Los sujetos perdidos durante el seguimiento pueden haber tenido resultados distintos a los disponibles para la evaluación?

¿En una cohorte abierta o dinámica? (8)

Crterios para la lectura crtica de artculos de diagnsticos

- ¿Existi3 una comparaci3n con una prueba de referencia adecuada? ¿Es correcto el patr3n de oro?
- ¿Incluy3 la muestra un espectro adecuado de pacientes? ¿C3mo se seleccionaron?
- ¿Existe una adecuada descripci3n de la prueba? Defini3n de un resultado positivo y qu3 es un resultado negativo.
- ¿Reproducibilidad de la prueba?
- ¿Hubo evaluaci3n "ciega" de los resultados? (9).

Crterios para la lectura crtica de artculos de ensayos clnicos

- ¿Fue aleatoria la asignaci3n de los pacientes a los tratamientos?
- ¿Se mantuvo oculta la secuencia de aleatorizaci3n?
- ¿Fueron todos los pacientes adecuadamente considerados hasta el final?
- ¿El seguimiento fue completo?
- ¿Se interrumpi3 precozmente el estudio?
- ¿Se analizaron los pacientes en el grupo al que fueron aleatoriamente asignados?
- ¿Se mantuvieron ciegos al tratamiento los pacientes, los clnicos y el personal del estudio?
- ¿Fueron similares los grupos al comienzo del ensayo?
- ¿Al margen de la intervenci3n en estudio los grupos fueron tratados de igual modo? (10).

Crterios para la lectura crtica de artculos de revisiones (revisiones sistemáticas y metaanálisis)

- ¿Los resultados de la revisi3n son válidos?
- ¿Se hizo la revisi3n sobre un tema claramente definido?
- ¿Buscaron los autores el tipo de artculos adecuado?
- ¿Crees que estaban incluidos los estudios importantes y pertinentes?
- ¿Qu3 bases de datos bibliogríficas se han usado?
- Seguimiento de las referencias.
- Contacto personal con expertos.
- Búsqueda de estudios no publicados.
- Búsqueda de estudios en idiomas distintos del ingl3s.
- ¿Crees que los autores de la revisi3n han hecho suficiente esfuerzo para valorar la calidad de los estudios incluidos?
- ¿Son las metodologías de los estudios y sus variables resultado adecuadas para realizar un metaanálisis?
- ¿Se ha realizado un test de heterogeneidad en los metaanálisis?
- ¿Se ha realizado test de subgrupos en los metaanálisis? (11).

Crterios para la lectura crtica de artculos de reglas de predicci3n clnica

- ¿Se define claramente el tipo de pacientes a los que se les aplica la regla?
 - ¿Están adecuadamente descritas las variables predictoras?
 - ¿El desenlace es relevante y tiene sentido clnico?
 - ¿La poblaci3n a estudio de la que se deriv3 la regla, incluy3 un espectro adecuado de pacientes?
 - ¿Se valid3 la regla en un grupo diferente de pacientes?
 - ¿Hubo una evaluaci3n ciega del desenlace y de las variables predictoras?
 - ¿Se midieron las variables predictoras y el desenlace en todos los pacientes?
 - ¿Se describen los métodos de derivaci3n y validaci3n de la regla?
 - ¿Se incluyen las variables importantes y los criterios de positividad?
 - ¿Se describe la reproducibilidad de las medidas? (12)
- En la tabla I se reflejan otras guías que podemos utilizar para la valoraci3n e interpretaci3n de los distintos tipos de artculos científicos.
- Otra forma de valorar si un artculo es de calidad consta en fijarse en la propia publicaci3n y el hecho de que esta revista disponga de revisiones por pares, que nos indicar3 si el artculo ha sido evaluado por revisores. Adem3s, el título nos dar3 una pista muy importante acerca del tema, y el hecho que venga la poblaci3n o el tipo de diseño utilizado podrían ser indicadores de la calidad del mismo. En el resumen tienen que aparecer todas las partes del artculo tal y como recomienda el "International Committee of Medical Journal Editors"; objetivo, diseño, número y principales características de los sujetos incluidos, tratamiento, principales resultados y conclusiones (19,20).

En una segunda aproximaci3n se analizar3 en profundidad la metodología, clave para determinar la capacidad del artculo para poder contestar adecuadamente, y de forma fiable, a nuestra pregunta.

En una tercera aproximaci3n se realiza un análisis más meticuloso sobre el material y métodos (19).

EL PAPEL DE LA ESTADÍSTICA EN LOS ARTÍCULOS CIENTÍFICOS

La bioestadística, es una valiosa herramienta que nos permite validar las conclusiones del trabajo científcico. En general, los autores presentan los resultados de sus estudios mediante gráficos, cuadros y tablas con datos cuantitativos, (medias, errores estándar, medianas) mencionando normalmente las pruebas estadísticas realizadas.

Miles de artculos científcicos se publican anualmente en cientos de revistas dedicadas a la difusi3n de búsquedas procedentes de cirugía.

TABLA I

PRINCIPALES GUÍAS DE LOS DIFERENTES TIPOS DE ESTUDIOS

<i>Tipo de estudio</i>	<i>Guías/checklist</i>	<i>Acrónimo</i>	<i>Referencia</i>
Ensayos Clínicos Aleatorizados	Escala de Jadad	Jadad	Jadad et al (13)
Metaanálisis	The Quality of Reporting of Meta-analyses (QUOROM)	QUOROM	Moher et al. (14)
	Meta-analysis of Observational Studies in Epidemiology (MOOSE)	MOOSE	Stroup DF et al. (15)
Revisiones sistemáticas	Cochrane Reviewers' Handbook Guides to the Medical Literature		Cochrane Group JAMA Group (16,17)
Guías de práctica clínica	Appraisal of Guidelines Research and Evaluation	AGREE	AGREEE project (18)

Si navegamos por las páginas de las revistas médicas podemos encontrar de forma insistente la presencia de $p < 0,05$, o $p > 0,05$, lo que hace referencia a la significación estadística en un contraste de hipótesis. Es en la búsqueda de una "p" en lo que el autor fundamenta toda la importancia de su estudio y el hecho de que considere brillante y/o concluyente el fenómeno o hecho estudiado.

Al parecer, este fenómeno se ha agravado con la llegada de los ordenadores, lo que facilita considerablemente los cálculos. Sin embargo, no todos los autores están dispuestos a usarlo correctamente.

Como ejemplo pondremos los cuidados postoperatorios tras una cirugía con anestesia general. Tenemos dos grupos de estudio uno a la planta y el otro grupo a la unidad de cuidados intensivos. El objetivo de este estudio fue evaluar la depresión respiratoria como posible complicación entre los grupos.

El resultado fue presentado con $p > 0,05$, es decir, ninguna diferencia estadísticamente significativa entre los grupos. Los autores, basándose en esto, concluyeron que los pacientes no corren el riesgo de depresión respiratoria. Sin embargo, el valor de $p > 0,05$, "una diferencia no estadísticamente significativa", sugiere falta de pruebas de un efecto. Cuando uno lee "no hubo diferencia estadística entre los dos grupos" se encuentra frente a información incompleta. El alto valor de "p" no significa estrictamente falta de efecto como los autores erróneamente concluyeron, sino que los datos eran insuficientes para establecer la necesidad de mantener la vigilancia en el postoperatorio de estos pacientes. Otros artículos omiten la palabra "estadísticamente" y escriben: "No hubo diferencias entre los grupos" o "diferencia significativa entre los grupos." Las diferencias pueden ser importantes desde un punto de vista clínico. En este ejemplo, sólo un paciente presentó depresión respiratoria requiriendo ventilación. Esto desde el punto de vista clínico es muy importante.

Como no es práctico estudiar todos los individuos afectados, se obtiene una muestra representativa de la población. Podemos explicar una $p > 0,05$ simplemente

porque el número de individuos estudiados es muy pequeño.

La probabilidad de que el estudio tenga la capacidad de detectar el fenómeno estudiado cuando existe se llama "potencia". La potencia de un estudio depende de la variabilidad del grupo, del tamaño muestral, del fenómeno observado y del nivel de significación. Sería razonable pensar que la depresión respiratoria como complicación postoperatoria, que no se manifestó en 30 pacientes, pudiera pasar en el paciente 32 si la muestra fuera de 35 pacientes.

Por otro lado hay que tener cuidado a la hora de interpretar los resultados en las grandes muestras. Es un error muy común en los estudios científicos donde se puede llegar a pensar que las muestras grandes, por ejemplo 5.000 casos, puedan inferir los resultados a la población de estudio.

Por lo tanto, el autor debe, antes de iniciar el estudio, planificar cuidadosamente el tamaño de la muestra, para asegurarse de que es apropiado para sus objetivos y por lo tanto siempre es un criterio de calidad de un artículo que se especifique como se ha realizado el cálculo del tamaño muestral, indicando que los autores han seguido una metodología rigurosa (21).

Es necesario observar si el autor describe detenidamente el modelo utilizado (y por qué) o simplemente se refiere a un método automático de selección de variables. No es suficiente hablar de los parámetros que se utilizaron sin garantizar que fueron asignados correctamente.

Un solo artículo no es suficiente para tomar una decisión sobre un fenómeno. Es muy común encontrar varios estudios sobre el mismo tema con diferentes conclusiones. Como se ha mencionado antes, los valores de $p > 0,05$ no garantizan la equivalencia, pero indica la falta de evidencia de una diferencia significativa.

Una comparación entre los estudios podría ser errónea por intentar comparar estudios que no son apropiados, cuya metodología no fue debidamente planificada. Los estudios multicéntricos que combinan los datos de diferentes lugares, son los más fiables (22).

