

CAPÍTULO 12.

LECTURA CRÍTICA DE UN ESTUDIO.

Javier Ballesteros

Departamento de Neurociencias – Psiquiatría.
Universidad del País Vasco, UPV-EHU.

1. Introducción.

Este capítulo sirve de lectura complementaria al capítulo sobre los criterios de publicación establecidos para diversos diseños de investigación clínica. Allí se presentan tanto los apartados que necesariamente debieran estar recogidos en la publicación de resultados según los distintos tipos de diseños, como la representación gráfica del desarrollo del estudio, o flujograma. Esos criterios, y diagramas de flujo, son una ayuda fundamental para poder valorar la calidad metodológica (validez interna) de un determinado estudio. Sin embargo, puede que no sean suficientes, o que se recojan de manera incompleta si el estudio fue publicado con antelación a la adopción de dichos criterios por las principales revistas científicas. Así el objetivo de este capítulo es orientar al lector sobre cuál es la información que debe perseguir – y localizar - en un artículo científico, dónde puede hallarla en el texto, y presentar algunos aspectos generales sobre su valoración. Valoración, que necesariamente va a tener un carácter general respecto de los objetivos perseguidos por el estudio. Plantillas específicas para valorar la calidad metodológica y los resultados según los objetivos de los diversos estudios de investigación pueden consultarse – entre otras – en la página web del Centro de Medicina Basada en la Evidencia de Oxford (<http://www.cebm.net>), y serían un buen añadido a las ideas que quiero expresar.

2. De la lectura general a la lectura crítica de un artículo científico.

El día sólo tiene 24 horas, y el lapso de vida de cada uno de nosotros es limitado, aunque no sepamos cuanto de limitado puede ser. Dicho de otra manera, y como profesionales en formación continuada a lo largo de todo nuestro desarrollo y actividad profesional, el tiempo que podemos dedicar a estar al día en nuestro ámbito de conocimiento es limitado, y mucho más teniendo en cuenta que el ámbito de nuestras lecturas de estudios científicos normalmente desbordará el estrecho margen de nuestra área. Así, una de las preguntas básicas en la formación continuada es la de ¿cómo podemos optimizar nuestro tiempo en la lectura de artículos científicos?

Un famoso algoritmo de lectura publicado cuando el autor era aún estudiante de Medicina, sugería una estrategia cuyos primeros pasos eran los siguientes. Primero lea el título del artículo, si no le interesa deséchelo, si le interesa pase al segundo paso. No es un mal comienzo, y seguro que todos lo suscribiríamos. En el segundo paso se planteaba si se conocía a los autores o estos procedían de un centro “respetable”, si era así se continuaba con la lectura, y si no lo era se desechaba. Aquí ya empiezan a ir las cosas mal. El algoritmo continuaba con la lectura del resumen, y luego ya con la lectura de las otras partes del artículo, priorizando los aspectos narrativos (introducción, discusión) sobre los aspectos metodológicos y empíricos (métodos, resultados). Supongo que todos hemos leído artículos de esa manera – sobretodo los que no nos interesaban especialmente – pero supongo también que el lector habrá apreciado – cuando menos – un potente sesgo de selección en ese curioso modelo de lectura, que ciertamente ahorra tiempo – y mucho – pero que es muy discutible bajo el punto de vista del conocimiento adquirido.

Claramente, el artículo científico normal presenta una serie de apartados que tienen unos objetivos específicos. La introducción repasa someramente el estado de la cuestión y enmarca el trabajo en ella; los métodos describen cómo se ha realizado el estudio, qué instrumentos de medición se han utilizado, y cómo se han analizado los datos, y su objetivo es que cualquier otro investigador con conocimiento suficiente sea capaz de replicar el estudio; los resultados describen los hallazgos objetivamente, y sin hacer interpretaciones, que es el objetivo del apartado de discusión, dónde no sólo se hacen esas interpretaciones en comparación con lo que se conocía previamente al estudio, sino que se resumen sus principales hallazgos, se discuten sus limitaciones, y se plantean líneas posteriores de investigación.

De todo ello ¿qué es lo que interesa especialmente para una lectura crítica del artículo? Fundamentalmente dos apartados y un párrafo. El último párrafo de la introducción que es el que generalmente señala los objetivos o las hipótesis del estudio; el apartado de métodos que es el que permite valorar si el diseño del estudio es válido o no para responder a los objetivos planteados; y el apartado de resultados que va a permitir comprobar si las hipótesis que guiaron el estudio son o no refutadas. El resto del capítulo intenta presentar estas ideas de una manera más operativa, y aunque está guiado básicamente por la lectura crítica de ensayos clínicos aleatorizados (ECA's), creo que puede aplicarse con pocas modificaciones a los otros diseños causales discutidos en esta monografía.

3. Aspectos principales a valorar en un estudio científico.

Los estudios clínicos pueden clasificarse según los objetivos clínicos que persiguen. Así, si el objetivo es etiológico, los diseños más apropiados serán los de investigación básica, y los estudios observacionales de cohortes, y de casos y de controles. Si el objetivo es diagnóstico los estudios transversales son los apropiados. Si el objetivo es valorar la eficacia de un tratamiento, claramente los ECA's destacan sobre los demás, mientras que si el objetivo es el de valorar factores pronósticos tanto ECA's con tiempos de seguimiento apropiados, como estudios de cohortes serían los adecuados. Por lo tanto, el primer aspecto a valorar en un estudio es cuál es su objetivo (normalmente expresado en el último párrafo de la introducción), y si el diseño del estudio es el adecuado para ese objetivo (primer apartado de la sección de método).

Otros aspectos a chequear en la sección de método son si ha habido o no una estimación previa de los tamaños de muestra a alcanzar; si esa estimación se ha realizado con criterios

realistas y en base a la variable principal de resultado, si la información de esa y otras variables se ha recogido con instrumentos o escalas apropiados, y que presenten una adecuada validez y fiabilidad; si ha habido enmascaramiento de los tratamientos en el caso de que se pudiera hacer, o al menos enmascaramiento del evaluador; si el seguimiento se ha producido de la misma manera para los grupos en comparación a lo largo de todo el estudio; y si los métodos de análisis estadístico propuestos están de acuerdo con el diseño del estudio y con el tipo de variable principal de resultado.

En la sección de resultados, hay varias piezas de información básicas. En primer lugar la valoración de la similitud de los grupos en estudio, lo que normalmente se recogerá en la tabla 1 de resultados y/o en el texto. Si se trata de un ECA y la asignación aleatoria se ha realizado de manera correcta, o si no hay grandes sesgos de selección en estudios observacionales, debería de haber una similar distribución de variables clínicas, y de pronóstico, en línea de base entre los grupos en comparación. Si esto no es así, algo ha ido mal, sea en el proceso de aleatorización o en el de formación de los grupos en comparación. Dado que en los estudios prospectivos siempre hay pérdida de sujetos, es importante valorar como se han introducido estas pérdidas en los análisis. Pueden no haberse introducido, en cuyo caso estaríamos en la situación – no aconsejable – de análisis por protocolo, en la que sólo analizamos los datos de los sujetos que han llegado al final del estudio. Si los sujetos que se han perdido lo han hecho de manera totalmente aleatoria, no habría en principio demasiados problemas en realizar este tipo de análisis. Pero la situación normal es la de pérdida diferencial de sujetos según ramas de tratamiento, y por mecanismos que no tienen nada de aleatorios. En esta situación los análisis basados en la cumplimentación del protocolo suelen estar sesgados, y es más lógico analizar a todos los sujetos incluidos en el estudio de acuerdo a la rama de tratamiento en la que empezaron, independientemente de que hayan salido o no del estudio durante el seguimiento. Esto es lo que se denomina análisis por intención de tratar. El problema es cómo estimar la información de aquellos sujetos que se han perdido. Esto suele realizarse de dos posibles maneras. Una mediante lo que se denomina arrastrar la última observación válida del sujeto hacia adelante, es decir asumir que el sujeto perdido presentaría en las evaluaciones en las que no está, al menos la última puntuación recogida. Esta es una opción relativamente conservadora – asume que el sujeto no modificaría su puntuación en las escalas apropiadas una vez que ha salido del estudio – pero la más recomendable. La otra opción, u otro grupo de opciones, es la estimación de cuáles hubieran podido ser sus puntuaciones en base a su similitud con otros sujetos de los que si tenemos puntuaciones completas a lo largo de todo el estudio. Este proceso de imputación es – cuando menos – arriesgado y no suele estar basado en los posibles mecanismos de pérdida subyacentes. Dos últimos aspectos de la sección de resultados son también importantes. En primer lugar, el valorar si los análisis estadísticos se basan en los resultados primarios apropiados, y en segundo si esos resultados dan la información adecuada. Normalmente, en cualquier diseño de investigación hay unos resultados primarios que se derivan implícitamente de las hipótesis u objetivos formulados, y todo un conjunto de resultados secundarios accesorios. La adecuada interpretación de los resultados siempre debe basarse en los resultados primarios y en los análisis especificados en el protocolo del estudio. Habría que sospechar que algo pasa (normalmente la caída y captura de valores $p < 0.05$) cuando el interés se desplaza hacia resultados secundarios, o hacia el análisis de subgrupos que no estaban claramente especificados en el protocolo del estudio. Por otro lado, los resultados estadísticos deben dar la información adecuada y esta

no es únicamente el valor p asociado a la hipótesis nula. Una significación estadística puede conseguirse de diversas maneras (la más lógica aumentando el tamaño de muestra) pero ello no significa que necesariamente tenga una significación clínica. La presentación de resultados mediante estimaciones adecuadas del tamaño del efecto (diferencia de medias, odds ratios, etc...), junto con la presentación de su precisión (normalmente intervalos de confianza al 95% para la estimación puntual) van ganando terreno a la presentación única de valores p. Sin embargo en ciertas áreas, y en ciertas revistas, los valores p aún sigue siendo el estándar en la presentación de resultados, aunque no permitan interpretar adecuadamente la significación clínica de las diferencias sobre las que se basan.

Todos estos aspectos hacen referencia fundamentalmente a la validez interna de los resultados de un estudio. Otro aspecto a tener en cuenta es su validez externa, la posible extrapolación de sus resultados a otros medios distintos de aquellos en los que el estudio se realizó, y para ello el mejor criterio es el de la replicación de sus resultados en diferentes medios. Hasta cierto punto, el metaanálisis basado en revisiones sistemáticas sin sesgos de publicación, pudiera apuntar en ese sentido dado que uno de sus objetivos es valorar la homogeneidad de los resultados que se combinan, y llegar a una estimación combinada que puede interpretarse como un indicador de la replicabilidad de una asociación causal. Los últimos aspectos que deben de tenerse en cuenta a la hora de valorar críticamente los resultados de un artículo reflejan el estudio de seres humanos y la declaración de conflicto de intereses. Todo estudio en el que participen seres humanos debe contar con la aprobación de su protocolo por un CEIC autorizado (Comisión de Ética e Investigación clínica), y la garantía de que ha habido un consentimiento informado por parte de los sujetos que han participado en el estudio. Sabemos que los estudios patrocinados por las empresas farmacéuticas suelen sobreestimar la eficacia de sus tratamientos con respecto a los resultados que se reportan en estudios independientes posteriores, y que determinados conflictos de intereses suelen asociarse también a la presentación de resultados positivos. Por todo ello, es necesario que cualquier estudio clínico presente de una manera clara tanto su financiación como los posibles conflictos de intereses que puedan tener sus autores.

4. Guía práctica para la evaluación de artículos sobre tratamiento o pronóstico.

Este apartado final presenta ya de manera resumida los aspectos que considero importantes a la hora de evaluar críticamente un artículo de investigación clínica. Estoy seguro de que no es una guía completa, ni de hecho lo pretende ser, pero sí una orientación general que se basa en otras guías¹ y recursos ya comentados (www.cebm.net).

1. Aspectos generales

- La evaluación crítica de un artículo clínico debe basarse en la lectura específica de los apartados de métodos y de resultados
- En ese sentido no nos interesa, en principio, ni el enmarque del estudio en el conocimiento actual (introducción), ni las opiniones de los autores sobre la relevancia de sus aportaciones respecto de otras (discusión)

2. Sección de introducción

- Último párrafo: objetivos o hipótesis del estudio

3. Sección de métodos

- Diseño del estudio: ¿es capaz de responder adecuadamente a los objetivos o hipótesis planteados?

- Tamaño de la muestra: ¿se ha calculado?, ¿se ha calculado con asunciones realistas?
- En su caso, aleatorización: ¿está suficientemente descrita?
- Cegamiento: ¿se ha mantenido la comparabilidad de los grupos a lo largo del estudio?, ¿estaban los evaluadores cegados respecto de los grupos a los que los sujetos fueron asignados?
- Descripción de los resultados a analizar: ¿está suficientemente claro cuáles son las principales variables de resultado, y cuáles son las variables secundarias?
- Medición de variables: ¿es la medición de las variables, sobretodo de las variables de resultado, suficientemente válida y fiable?, en el caso de que los instrumentos de medición sean heteroadministrados ¿hay información sobre la concordancia de los evaluadores en las mediciones realizadas?
- Análisis estadístico: ¿son las pruebas estadísticas previstas apropiadas dado el diseño del estudio y las características de las variables de resultado?, ¿cómo se van a valorar las pérdidas durante el seguimiento?, ¿se van a presentar análisis por intención de tratar o análisis por cumplimentación de protocolo?

4. Sección de resultados.

- Transcurso del estudio: ¿se representa el flujo de sujetos en el estudio mediante un diagrama?, o ¿sería factible hacerlo en función de la información que se recoge en el texto, tablas y gráficos?
- Primera tabla: ¿son los grupos que se comparan suficientemente similares en la línea de base respecto de variables clínicas y de pronóstico?
- Resultados primarios: deben presentarse en texto, tabulados, o gráficamente, pero con el suficiente detalle como para que se puedan recalcular las pruebas de significación estadística, la estimación del efecto, y su precisión; tienen que estar en relación directa con las hipótesis del estudio, y las variables de resultado descritas
- Otros resultados: deben presentar fundamentalmente efectos adversos y pérdidas en el seguimiento. Desconfiar de los análisis por subgrupos (a no ser que así esté establecido en las hipótesis del estudio), y de todo análisis accesorio no directamente relacionado con el objetivo del estudio

5. Otros aspectos.

- Justificación de que el protocolo del estudio ha sido aprobado por un CEIC, y de que todos los sujetos han otorgado su consentimiento informado para participar en el estudio
- Presencia de información sobre la financiación del estudio, y sobre las declaraciones de intereses de sus autores

5. Referencias.

1. Lawrie SM, McIntosh AM, Rao S. Critical appraisal for psychiatry. Edinburgh: Churchill Livingstone; 2000