

## Documento

# GUÍA PARA EL ANÁLISIS CRÍTICO DE PUBLICACIONES CIENTÍFICAS

Dr. *Jorge A. Carvajal C. Ph.D.*

Unidad de Medicina Materno-Fetal, Departamento de Obstetricia y Ginecología, Facultad de Medicina, Pontificia Universidad Católica de Chile

## INTRODUCCIÓN

Antes de creer la información que un paper contiene, y modificar la práctica clínica en base a sus resultados, debemos verificar si la investigación es metodológicamente correcta. El objetivo de esta guía es recomendar algunos pasos a seguir para evaluar críticamente los resultados de una investigación, decidir sobre la validez de los resultados publicados, y aceptar su aplicación a la práctica clínica (1).

El análisis de una publicación científica se basa en tres etapas:

1. Definir la relevancia del tema estudiado en el trabajo:

- a. Relevancia Clínica: sobre qué trata el estudio.
- b. Escenario Clínico: respecto de que situación clínica específica versa el estudio.
- c. Pregunta: cuál es la pregunta que se intenta responder.

2. Descripción del estudio:

- a. Diseño.
- b. Pacientes.
- c. Intervención.
- d. Resultados.
- e. Conclusiones.

3. Análisis Crítico:

a. Validez Interna: calidad metodológica del trabajo.

b. Validez Externa: aplicabilidad.

c. Comentario: nuestras personales conclusiones basadas en los datos crudos mostrados por el estudio y en la evaluación metodológica efectuada.

En esta guía, la que en gran proporción es una adaptación de un artículo publicado (1), más apuntes personales del Curso de Investigación Clínica Controlada, dictado por el Centro Rosarino de Estudios Perinatales, presenta una detallada metodología, en 8 pasos, para efectuar el análisis crítico de los artículos publicados. Si bien seguimos estos 8 pasos para el análisis de los artículos, sólo parte del seguimiento de ellos estarán presentes en la redacción del análisis que presentaremos en cada número de la Revista Chilena de Obstetricia y Ginecología en la sección Revista de Revistas.

## PASO 1. Título del trabajo, autores e introducción

La evaluación detallada del título del artículo permite conocer la gran mayoría de las veces el diseño del estudio y sus conclusiones. No es fácil dar un título a un artículo, y es por lo tanto una preocupación de los autores el elegir un título que efectivamente represente su pensamiento y en alguna medida el diseño del artículo. Especial atención debemos poner porque como veremos adelante, muchas veces las conclusiones de los autores no son sustentadas por los datos o el diseño, y esto con frecuencia se reflejará en un título inapropiado.

La observación cuidadosa del o los autores nos informará de si el estudio es la consecuencia lógica de un estudio previo que el grupo ya venía desarrollando, y por cierto de la credibilidad de los datos cuando se trate de investigadores de presti-

gio. Recordemos que es imprescindible que la pregunta que el artículo intenta responder tenga sentido biológico, y ello muchas veces resulta del trabajo previo de los autores

En la introducción, los autores hacen una revisión de la información respecto del tema a investigar, e informan de los objetivos del trabajo. La lectura de la introducción es especialmente útil para el lector, en el sentido de informarse de aquello que ya se sabe respecto del tema. Hay que poner atención para descubrir si se ha omitido información importante, ya que muchas veces los autores ocultan información para hacer que sus datos parezcan más novedosos o evitar antecedentes que estén en contra de sus resultados.

### **PASO 2. Evaluación de la originalidad del estudio**

Solo una pequeña porción de la investigación médica es completamente novedosa, y del mismo modo, sólo una pequeña porción repite exactamente un trabajo previo. La gran mayoría de la investigación nos dirá, en el mejor de los casos, que una hipótesis determinada es probablemente un poco más correcta o un poco más incorrecta que lo que era antes de agregar esta nueva información. Por ello, es perfectamente válido llevar a cabo un estudio que no es completamente original. De hecho la ciencia del meta análisis se basa en la aparición de más literatura destinada a responder la misma pregunta de un modo similar.

De este modo, la pregunta práctica a realizar no es ¿alguien ha hecho antes un estudio similar?, lo correcto es preguntarse ¿agrega esta investigación un conocimiento nuevo al problema en estudio? Por ejemplo:

– ¿Se trata de un estudio más grande, más largo, o definitivamente más sustancial que los estudios previos?

– ¿Se trata de un estudio cuya metodología es mejor y más rigurosa que estudios previos?. Particularmente, ¿contiene alguna crítica específica a la metodología de estudios anteriores?

– El aporte de datos de este trabajo ¿agrega significancia al metanálisis de los estudios previos?

– ¿La población estudiada es diferente? ¿Observó el estudio diferentes edades, sexo, o grupo étnico comparado a los trabajos previos?

– ¿Es el problema clínico de importancia suficiente? ¿Hay suficientes dudas en el ambiente público o decisiones claves que tomar que justifiquen la necesidad de nueva evidencia?

### **PASO 3. Evaluación de la población estudiada**

Antes de determinar si el resultado se aplica a su población y a su práctica clínica, debe determinarse cuales fueron los pacientes incorporados o potencialmente incorporados al estudio. En base a los criterios de inclusión y exclusión hacer conclusiones respecto de la validez externa de los datos, ello permitirá concluir si los resultados serán útiles para nuestro trabajo. Así por ejemplo, si los criterios de inclusión son muy restrictivos, ya sabemos que el resultado benéfico de una droga será útil sólo para un porcentaje muy pequeño de los pacientes. Para efectuar el análisis debe responder las siguientes interrogantes.

– ¿Cómo fueron reclutados los pacientes? Esto es importante por que el modo de selección de los pacientes determina si hubo o no un sesgo de selección. Si hubo un sesgo de selección, entonces los resultados del estudio pueden ser motivados por las especiales características de la población reclutada. Ejemplo: si los pacientes son invitados al estudio mediante un aviso puesto en la cafetería del hospital, limitará la población sólo a personas de mejores ingresos y con determinados hábitos alimenticios.

– ¿Quién fue incluido en el estudio? Esta pregunta se relaciona con los criterios de selección, si estos son muy estrictos, la población no será representativa de todos los pacientes, y los resultados sólo serán aplicables a ese grupo. Si evaluamos la farmacología de una droga en varones sanos universitarios, los resultados no serán aplicables a personas de otro sexo o edad.

– ¿Quién fue excluido del estudio? Los criterios de exclusión son útiles para hacer el estudio más homogéneo, pero limitan la aplicabilidad de los resultados. Si se excluyen las formas leves de la enfermedad, el estudio efectuado con formas graves no será aplicable a casos leves-moderados. Así por ejemplo un estudio que incluyó sólo pacientes hospitalizados no tendrá conclusiones válidas para el manejo de pacientes ambulatorios.

– ¿Fueron los sujetos estudiados en circunstancias reales? Esta observación no invalida los resultados, pero señala que su efectividad en el manejo real debe ser evaluada. Por ejemplo: si se evalúa un medicamento, y se tiene a una persona que hace educación para que los pacientes tomen sus fármacos, los llama a la casa, les cuenta las pastillas, etc., es probable que la adherencia al tratamiento sea mucho mejor que la que se logrará en la práctica habitual.

#### PASO 4. Evaluación del diseño del estudio

Aunque el análisis del diseño del estudio pueda asustar, mucho de lo que conocemos como “apreciación crítica” es solamente sentido común. Debemos en primer lugar identificar el diseño del estudio. Esta identificación permite saber cual es el tipo de conclusiones a que el estudio puede conducir. Por ejemplo si el estudio es un diseño de caso control, las conclusiones pueden llevar a asociación estadística de riesgo, pero no a conclusiones de causalidad. Si el diseño es de un estudio prospectivo doble ciego randomizado, se sugiere seguir las instrucciones contenidas en el Consort Statement (2) para verificar que el diseño contenga todos los elementos necesarios para que el estudio se lleve a cabo. El análisis debe continuar luego con dos preguntas fundamentales:

– ¿Qué intervención o maniobra está siendo comparada?, y ¿contra cual intervención se está comparando? Existe la tentación a creer lo publicado como lo que efectivamente sucedió. Sin embargo, muchas veces los autores (de modo consciente o inconsciente) efectúan un arreglo “cosmético” de sus decisiones y manejo, sobre estimando la originalidad, rigurosidad y potencial importancia del estudio. Para evitar este problema debemos: 1. Evaluar el método de obtención de los datos, para determinar si efectivamente es capaz de obtener con acuciosidad los datos deseados (ejemplo: análisis de fichas, las cuales no necesariamente son 100% confiables); 2. Evaluar si hay suficiente información respecto de la intervención efectuada (dosis del medicamento y naturaleza del placebo; características detalladas de la intervención) y de los sujetos reclutados (cuantos fueron invitados a participar, cuantos realmente aceptaron, hay balance entre razas y sexos, etc.); 3. Evaluar si se controlaron factores confundentes (evaluar si los pacientes del grupo estudio y control difieren solo en la intervención estudiada y no en otra intervención adicional); 4. Evaluar si la información es realmente original y contributiva (significa si ya hay muchos estudios similares).

– ¿Qué resultado fue medido y como se midió?. Debe evaluarse si el resultado medido es de importancia con relación al problema que suscita el estudio. De este modo si se prueba un nuevo tratamiento para una enfermedad potencialmente mortal, el resultado clínicamente relevante a evaluar es el tiempo de sobrevida y la calidad de vida. No es de interés evaluar el nivel plasmático de alguna sustancia que potencialmente se correlaciona con la posibilidad de una mayor sobrevida.

La medición de un efecto sintomático (dolor), funcional (movilidad), psicológico (ansiedad) o social (incomodidad social) está lleno de dificultades. En estos casos debe buscarse evidencia en el artículo de que el resultado medido fue objetivamente validado; esto es por ejemplo confirmar que alguien ha validado la escala de dolor o ansiedad utilizada, y que esta escala realmente mide lo que se desea que mida; y finalmente que cambios en la medida del resultado reflejen verdaderos cambios en el estado del paciente, esto porque cambios importantes para el médico no necesariamente lo son para el paciente o viceversa.

#### PASO 5. Evaluación de los posibles sesgos

El sesgo sistemático se define como cualquier cosa que influya en las conclusiones acerca de los grupos y distorsione las comparaciones. Ya sea que el estudio sea randomizado y controlado, o un estudio comparativo no randomizado, o de cohorte, o caso-control, el objetivo es que los grupos sean comparables excepto respecto de la particular diferencia que se está estudiando. Así por ejemplo, los pacientes deben, tanto como sea posible, recibir las mismas explicaciones, tener el mismo contacto con los médicos, ser evaluados el mismo número de veces y por los mismos resultados. Diferentes diseños tienen diferentes necesidades para evitar el sesgo sistemático.

*Estudios randomizados controlados:* en un estudio randomizado el sesgo sistemático, en teoría, es evitado por seleccionar pacientes de una población dada y asignarlos al azar a los diferentes grupos. Sin embargo, potenciales fuentes de sesgo sistemático en este tipo de estudios son:

– Sesgo de selección: el cual se origina por una randomización inadecuada, haciendo que los grupos tengan diferencias.

– Sesgo de intervención: se origina si los pacientes son manejados de un modo diferente según el grupo, de ese modo habrá más de una intervención diferente sobre cada grupo.

– Sesgo de exclusión: se origina si la pérdida de pacientes es mayor en uno de los dos grupos, principalmente si se pone menos atención al grupo control.

– Sesgo de detección: si la medición de los resultados se efectúa de modo diferente, por ejemplo utilizando un método más sensible en el grupo estudio.

*Estudios clínicos controlados no randomizados:* Estos estudios agregan a los sesgos descritos para los estudios randomizados, el sesgo poten-

cial de la falta de randomización. No rara vez los pacientes son asignados a un grupo mediante eventos que significan potencialmente un sesgo. Por ejemplo: si los pacientes dados de alta del hospital son asignados a un tratamiento, mientras que los en control en policlínico son asignados a otro tratamiento.

Como regla general, si el estudio es no randomizado, se debe usar el sentido común para evaluar si las diferencias basales entre el grupo estudio y el grupo control son suficientemente grandes como para invalidar las diferencias detectadas y adscribir a la intervención en estudio. Con mucha frecuencia este es el caso en este tipo de diseños.

*Estudios de cohorte:* La selección de un grupo comparable es el problema más difícil en este tipo de estudios. Solo pocos estudios de cohorte logran encontrar grupos que siendo equivalente en edad, sexo, estatus socioeconómico, enfermedades coexistentes, etc., difieren sólo en la intervención estudiada. En la práctica mucho del control de variables en este tipo de artículos se efectúa en la etapa de análisis, en que observamos complejas correcciones estadísticas para ajustar por la variabilidad basal de los grupos.

*Estudios caso-control:* En estudios caso-control (en el cual la experiencia de un individuo con una enfermedad particular es analizado de modo retrospectivo para identificar un agente potencialmente causante), el proceso más potencial de sesgo no es la medición del resultado, sino que el diagnóstico de "no caso" (control) y la decisión de cuando el individuo se transforma en un caso.

## PASO 6. Evaluación del doble ciego

Incluso los intentos más acuciosos para generar grupos comparables serán perdidos, especialmente si la persona que mide el resultado sabe a que grupo pertenece el paciente. El investigador podría reevaluar el resultado si este es inapropiado al tratamiento que se está evaluando. Si por ejemplo se está probando una droga hipotensora, y la paciente es del grupo estudio, pero la presión arterial está más alta, el evaluador tenderá a medir nuevamente la presión.

Es importante que se haga el máximo esfuerzo posible por lograr el "ciego" tanto al paciente como al médico. Esto, muchas veces requiere, que el investigador no participe directamente en la atención de los pacientes. Si el estudio fue metodológicamente bien logrado, y los investigadores pusieron especial énfasis en lograr el enmascaramiento de la intervención, ello debe

aparecer reflejado en la sección de métodos en detalle, y no como una simple frase.

## PASO 7. Evaluación de los métodos estadísticos utilizados

Para lograr este paso no se requiere ser un experto en estadística, pero si se requiere tener algunos conceptos claros sobre como y cuando aplicar un determinado tipo de test estadístico. Existen tres elementos estadísticos a los que se debe poner especial atención, y los cuales deben estar explícitamente detallados en la sección de métodos del paper:

1. El cálculo del tamaño muestral.
2. La duración estimada del seguimiento y el número estimado de pacientes que lo completarán.
3. El análisis según intención de tratar.

1. *Tamaño muestral:* Un estudio debe ser suficientemente grande para tener una alta probabilidad de detectar con significancia estadística un efecto benéfico si es que existe, y estar razonablemente seguro de que el beneficio no existe si el estudio no lo detecta.

Para calcular el tamaño muestral, el clínico debe decidir dos cosas. La primera decisión es cual nivel de diferencia entre ambos grupos constituirá un efecto clínicamente significativo. Este valor es por cierto diferente de la diferencia estadística. En segundo lugar deberá estimar el promedio y desviación estándar de la principal variable medida.

Mediante el uso de un nomograma los autores podrán, antes de efectuar el estudio, saber el tamaño muestral necesario para tener una probabilidad moderada, alta o muy alta de detectar una diferencia que realmente existe entre los dos grupos, esto es lo que se denomina poder estadístico, y habitualmente se fija entre 80-90%. Existen numerosos estudios publicados con bajo poder, probablemente porque para los autores resultó mucho más difícil de lo estimado el reclutar a los pacientes. Esto lleva al error de tipo II o  $\beta$ , que es la conclusión errónea de que la intervención no tuvo efecto, que es lo opuesto del error de tipo I o  $\alpha$ , que es más raro y corresponde a la conclusión de que existe una diferencia que no existe.

2. *Duración y lo completo del seguimiento:* Aunque el tamaño muestral sea suficientemente grande, el estudio debe continuar por un tiempo necesariamente largo como para que el resultado principal sea evaluado. La duración depende exclusivamente de la variable que se desee analizar.

Los sujetos que se retiran del seguimiento comparado con los que siguen el estudio, con mayor probabilidad no tomaron los medicamentos como se les indicó, con más frecuencia perdieron sus controles médicos, o tienen mayor frecuencia de efectos adversos del medicamento. Las razones para que un paciente se retire del seguimiento son:

- Entrada incorrecta del paciente al estudio, esto es, que el investigador descubra durante el estudio que un paciente fue randomizado y entró al estudio, pero no cumplía con los criterios de inclusión.

- Sospecha de reacción adversa a la droga en estudio. Hay que notar que la tasa de reacciones adversas al medicamento debe compararse con la tasa de reacciones adversas en el grupo placebo; las tabletas inertes son capaces de producir un rush alérgico en un número no despreciable de personas.

- Pérdida de la motivación del paciente.

- Retiro del estudio por decisión de su médico tratante. Esto sucede con frecuencia por la aparición de una enfermedad concomitante o por la sensación de falta de mejoría.

- Pérdida del seguimiento. Ejemplo: si la paciente se cambia de domicilio.

- Muerte del paciente.

3. *Análisis según intención de tratar*: El ignorar a todos aquellos que se retiren del estudio producirá un sesgo del resultado (sesgo de análisis), usualmente a favor de la intervención analizada (porque como se indicó aquellos con efectos adversos o mal resultado con mayor frecuencia se retiran del estudio). Por eso el estándar actual es analizar los resultados de estudios comparativos según “intención de tratar”. Esto significa que todos los datos de los pacientes randomizados a una rama del estudio deben ser utilizados para el análisis, es decir todo paciente que se randomiza debe considerarse en aquella rama en que fue randomizado. Esto incluye aquellos quienes dejaron el estudio antes de ser terminado, aquellos que no se tomaron las tabletas, e incluso aquellos que posteriormente recibieron la intervención control por cualquier motivo. Del mismo modo aquellos que fueron randomizados a placebo deben ser analizados con ese grupo, ya sea que tomaron o no el placebo, y si completaron o no el seguimiento. Las comparaciones entre aquellos que efectivamente tomaron las tabletas o recibieron la intervención completa, son sólo útiles como análisis secundarios que motiven nuevos estudios, pero

no son útiles para obtener conclusiones científicamente validadas. Así por ejemplo, en pocas situaciones no se utiliza el análisis según “intención de tratar”; el ejemplo más común son los estudios para análisis de eficacia, el que se utiliza para evaluar los efectos de la intervención en si misma, y por lo tanto analiza según la intervención efectivamente recibida. Pero, aun si los sujetos para un análisis de eficacia provienen de un estudio randomizado, para los propósitos del análisis ellos constituyen un estudio de cohorte.

## **PASO 8. Evaluación de los resultados y de las conclusiones**

Los resultados son la parte más importante del trabajo. Habitualmente los datos más relevantes aparecen presentados en tablas o gráficos. Estos deben ser observados con cautela para ver los datos crudos. Muchas veces los autores presentan los datos como porcentaje de un valor, para intentar hacer mayores pequeñas diferencias. Al final de los resultados el lector debe establecer sus propias conclusiones. Son estas las que realmente valen. La idea es que en este punto ya se conoce el problema a estudiar (introducción), el método usado (diseño experimental), y los resultados, por lo tanto es posible hacer las conclusiones propias, la que serán adecuadas a los datos y no estarán sesgadas por el deseo de determinado resultado, como ocurre a los autores.

No debe leer las conclusiones sin antes tener las propias, extraídas de los datos presentados. En las conclusiones los autores presentan un resumen de los datos y la interpretación de los mismo, las que no rara vez difieren de las propias. Las conclusiones válidas son sólo aquellas que se basan en los datos presentados, y no en tendencias, o en “grandes diferencias pero que no alcanzaron significación estadística”. Las conclusiones propias del lector orientan a lo que hará con la información obtenida, y no rara vez son diferentes de las de los autores.

## **RESUMEN**

1. Evaluar cuidadosamente el título, los autores y la introducción.
2. Determinar la contribución del estudio al conocimiento (originalidad)
3. Comprobar la población estudiada (validez externa del estudio).
4. Caracterizar el diseño.
5. Evaluar si se evitó el sesgo.

6. Comprobar si se enmascaró la intervención (doble ciego).

7. Evaluar si los detalles estadísticos fueron cuidados.

8. Leer los datos crudos y obtener conclusiones.

Si un estudio no cumple con todas las características deseadas para ser considerado metodológicamente impecable, no significa que deba ser descartado o eliminado. Lo importante es descubrir cuales son los problemas del estudio, y estimar cuales son las conclusiones a que validamente se puede aspirar considerando las falencias y características del diseño. El lector deberá por lo tanto

extraer sus propias conclusiones respecto de los resultados entregados (los datos son objetivos, la interpretación de los mismos no lo son), las cuales no rara vez difieren de las conclusiones de los autores.

#### **BIBLIOGRAFÍA**

1. Greenhalgh T. Assessing the methodological quality of published papers. *BMJ* 1997; 315(7103): 305-8.
2. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of parallel-group randomised trials. *Lancet* 2001; 357: 1191-94.