

¿Cómo leer la literatura científica? Analizando la evidencia que evalúa intervenciones terapéuticas. (PARTE V).

How to read a scientific article?
Randomized controlled trials. (PART 5)

Bottaro FJ

Servicio de Clínica Médica, Hospital Británico
Coordinador del Comité de Revisión Institucional, Hospital Británico

febottaro@hotmail.com

Fecha de recepción: 5/12/2014
Fecha de aprobación: 13/12/2014



COMO SE LEE

HEMATOLOGÍA Volumen 18 nº 3: 293-298
Septiembre-Diciembre 2014

Palabras clave: Aleatorización,
ensayo clínico controlado,
punto final,
factores pronósticos.

Keywords: Randomization,
controlled clinical trial,
outcome,
prognostic factors

"We think we understand the rules when we become adults, but what we really experience is a narrowing of the imagination" David Lynch

En esta entrega abordaremos inicialmente como evaluar la evidencia proveniente de estudios individuales que evalúan intervenciones terapéuticas. El mejor diseño para evaluar intervenciones terapéuticas son los "ensayos clínicos controlados" (ECC)⁽¹⁾. Sin embargo, a menos que sea un ECC aleatorizado de alta calidad y haya incluido un importante número de sujetos, no es el nivel de evidencia más alto para contestar una pregunta sobre una intervención terapéutica y lo ideal sería hallar una revisión sistemática. En nuestra práctica cotidiana, frecuentemente hallaremos un solo estudio individual para evaluar algunas intervenciones.

En los párrafos siguientes desarrollaremos aspectos claves del proceso formal de evaluación de estudios

que analizan intervenciones terapéuticas. Para el lector que quiera ampliar o profundizar este tema puede consultar dos excelentes capítulos que son la base del desarrollo de este manuscrito.^(2, 3)

Primera etapa

En primera instancia debemos decidir si el estudio que hemos seleccionado responde a nuestra pregunta de investigación⁽⁴⁾ y valdrá la pena leerlo (¿el objetivo del ECC se corresponde con la pregunta que me he planteado?). Con una rápida mirada al resumen del mismo podremos decidir si tendremos que continuar con nuestra búsqueda o leeremos el estudio seleccionado.

En este momento, también, puede ser conveniente

que analicemos algunas cuestiones que están asociadas con la validez externa del estudio (la posibilidad que los hallazgos del ensayo puedan ser replicados en un contexto diferente del contexto en el cual se desarrolló originalmente). Nos preguntaremos: ¿es la intervención estudiada aplicable en mi medio? ¿son los pacientes del estudio similares a mis pacientes? ¿es el punto final del ECC un evento de interés clínico que me permita valorar adecuadamente riesgos y beneficios de la intervención?

Una vez decidida la lectura del estudio procederemos a su análisis.

¿Son válidos los resultados de este estudio?

El abordaje usual para responder esta pregunta consiste en analizar el artículo seleccionado mediante tres simples preguntas compuestas que representan los criterios de validez para ECC. El esquema de tres preguntas se describe en la **Tabla 1**.

Tabla 1: Criterios de validez para la evaluación de artículos que evalúan intervenciones terapéuticas.

¿Son válidos los resultados de este estudio?

1) ¿Tienen ambas ramas de tratamiento (grupo intervención y grupo control) igual pronóstico al inicio del estudio?

- ¿El estudio fue aleatorizado? (cada sujeto que participa del estudio tiene igual chance de ser incluido en cualquiera de los dos grupos)
- ¿La asignación de los sujetos participantes al grupo de aleatorización fue realizada de manera apropiada?
- ¿Son similares ambos grupos al inicio del estudio? ¿los factores pronósticos más relevantes a considerar están balanceados entre ambos grupos control y tratamiento?

2) ¿Se mantuvo el balance de los factores pronósticos a lo largo del estudio?

- ¿El grupo al que fue asignado cada sujeto se mantuvo oculto (ciego)? ¿es un estudio abierto? ¿Quiénes fueron cegados: el paciente, el investigador y personal del estudio, el evaluador de puntos finales, todos?
- ¿Fueron analizados todos los sujetos participantes dentro de los grupos a los que fueron originalmente aleatorizados? (principio de intención de tratar)

3) ¿Los grupos se mantuvieron equilibrados al finalizar el estudio?

- ¿El seguimiento de los pacientes fue suficientemente largo y completo? ¿y similar en ambos grupos?
- ¿Se detuvo el estudio antes de tiempo?

¿Cuáles son los resultados del estudio?

- ¿Cuál es la magnitud del efecto del tratamiento?
- ¿Cuál es la precisión de los resultados del estudio?

¿Puedo aplicar los resultados de este estudio a mi paciente?

- ¿son los pacientes de mi práctica habitual similares a los sujetos que participaron del estudio?
- ¿fueron considerados en los resultados puntos finales de interés y relevancia clínica?
- ¿Son mayores los beneficios que los riesgos asociados a la implementación de la intervención en estudio?
- ¿Cuáles son los valores y preferencias de nuestros pacientes en relación tanto al evento que queremos prevenir y el tratamiento en estudio para tal fin?

¿Tienen ambas ramas de tratamiento (grupo intervención y grupo control) igual pronóstico al inicio del estudio?

1) *¿El estudio fue aleatorizado?*

El motivo por el cual los estudios en que la asignación al grupo tratamiento o al grupo control es elegido por preferencia del médico o del paciente (estudios observacionales) pueden llevarnos a conclusiones erróneas reside en que la mortalidad o morbilidad dependen de muchas causas; de las cuales la intervención en estudio es solo una de ellas. Estas “causas” se denominan factores pronósticos, y en los ECC se busca que estén balanceados entre las dos poblaciones (grupos control y tratamiento). Por ejemplo, en un ECC en el que se quiere evaluar el efecto de una nueva droga para la reducción del infarto agudo de miocardio, es esperable que la probabilidad de ocurrencia de un evento esté vinculada con más factores más allá de la administración o no de la droga en estudio (hipertensión, tabaquismo, dislipemia, etc.). Una manera de lograr el balance de factores pronósticos es la aleatorización. El poder de esta reside en que tanto el grupo control como el grupo tratamiento estarán balanceados tanto en los factores pronósticos que conocemos como en aquellos que aún no conocemos.

La literatura médica abunda de ejemplos en los que un ECC refuta resultados provenientes de estudios observacionales. Uno de los ejemplos más célebres es el de la terapia de reemplazo hormonal (TRH) para la prevención de eventos cardiovasculares en mujeres posmenopáusicas.

Durante mucho tiempo se sostuvo que la TRH en estas mujeres podía disminuir el riesgo de eventos coronarios (infarto de miocardio y muerte) y su uso fue ampliamente extendido en el mundo. Esta creencia médica estaba sostenida por plausibilidad biológica y por los resultados de múltiples estudios observacionales que así lo indicaban. Inclusive algunos estudios prospectivos de grandes dimensiones (cerca de 50000 pacientes)⁽⁵⁾ y meta-análisis de estudios observacionales⁽⁶⁾. Sin embargo, los resultados del primer ECC realizado en mujeres con enfermedad cardiovascular establecida tuvo resultados sorprendentes: la TRH falló en reducir los eventos coronarios⁽⁷⁾. Inclusive, el más reciente estudio proveniente de la serie Women Health Initiative, también demostró que la TRH no posee potencial preventivo

de eventos coronarios en mujeres posmenopáusicas sin antecedentes de enfermedad cardiovascular⁽⁸⁾. Otros ejemplos en la literatura médica en que los resultados de estudios observacionales son refutados por el ECC comprenden: vitaminas antioxidantes y cáncer gastrointestinal⁽⁹⁾, agentes antiarrítmicos (encainida y flecainida) y prevención de arritmias ventriculares en pacientes con infarto de miocardio⁽¹⁰⁾, milrinona oral e insuficiencia cardíaca⁽¹¹⁾, y muchos otros más.

¿Por qué existe esta diferencia entre los resultados de estudios observacionales y ECC? Como mencionamos anteriormente, en los estudios observacionales son las preferencias del médico y/o del paciente las que seleccionan si este recibirá o no el tratamiento. Frecuentemente, factores presentes como la presencia de comorbilidades, uso concomitante de otros tratamientos o drogas, las creencias del individuo, severidad de la enfermedad y sus síntomas pueden influir en el riesgo de desarrollar un evento. Si estos factores están distribuidos en forma asimétrica entre los grupos, el efecto de la intervención puede verse exagerado, suprimido o inclusive puede obtenerse resultados opuestos al esperado. En el ejemplo de la TRH en mujeres posmenopáusicas, posteriormente se halló que en los estudios observacionales el grupo tratamiento (TRH) estaba constituido por mujeres que tenían actividad física rutinaria, más controles médicos, y mejor estatus socioeconómico que las mujeres del grupo control, creando un desbalance de los factores pronósticos entre los grupos. Cualquiera sea los factores que desbalancearon los grupos (conocidos y no conocidos) hoy estamos seguros que fue el “pronóstico basal” en lugar de la TRH la explicación de los resultados de los estudios observacionales.

2) *¿La asignación de los sujetos participantes al grupo de aleatorización fue realizada de manera apropiada?*

No obstante, no solo es importante que el estudio sea aleatorizado, sino también la forma en que está se realizó (no profundizaremos en este aspecto en este capítulo) y la manera en que fueron asignados los pacientes a los grupos. Si la asignación de los pacientes al grupo que fueron aleatorizados es adecuada entonces el investigador no podrá predecir cuál será el tratamiento que le tocará al próximo paciente; y por lo tanto incapaz, consciente o inconscien-

temente, de distorsionar el balance entre los grupos que están siendo comparados. Veamos un ejemplo: frecuentemente el método de aleatorización es una secuencia generada al azar por computadora y la asignación de los grupos transcurre mediante el uso de sobres cerrados que contienen el grupo asignado. Si estos sobres permiten ver su interior ya que son traslúcidos (no opacos), o no están sellados permitiendo que el investigador pueda conocer el grupo de aleatorización antes que este sea asignado, entonces puede ocurrir que se manipule la asignación de pacientes. Un método de asignación de grupos inadecuado puede alterar el efecto de la intervención sobre el evento de interés; magnificándolo, suprimiéndolo o invirtiéndolo.

Muchas veces el método de asignación de grupos no está descrito en los manuscritos de ECC, pero podemos confiar en que si la aleatorización y asignación de grupos fue realizada por vía telefónica (ejemplo: IVRS, interactive voice response system), o mediante otro sistema que se encuentre a distancia del paciente e investigador, es muy posible que entonces este sea adecuado.

2) *¿Son similares ambos grupos al inicio del estudio? ¿los factores pronósticos más relevantes a considerar están balanceados entre ambos grupos control y tratamiento?*

Si la aleatorización y su asignación fueron adecuadas, los grupos que son comparados deberían estar adecuadamente balanceados. Es decir que los grupos deberían ser similares en la presencia de los factores pronósticos más conocidos o relevantes (excepto por la recepción de la intervención en estudio) al inicio del ensayo. Como dijimos anteriormente, el poder de la aleatorización reside en la distribución balanceada de los factores pronósticos conocidos, desconocidos y también de los potenciales confundidores. Sin embargo, a veces, a pesar que los participantes fueron aleatorizados, puede ocurrir que existen diferencias en la distribución de factores pronósticos. Si el método de aleatorización y asignación de grupos es adecuado, esto muy probablemente se deba al azar. Además debemos tener en cuenta que si el tamaño del estudio (número de pacientes que participan del mismo) es pequeño, más chances tendremos que los grupos queden desbalanceados. Cuanto mayor es el tamaño del estudio, más chances que los grupos estén adecuadamente balanceados.

Muchos investigadores al reportar las características basales de los participantes (habitualmente la **Tabla 1** de cualquier estudio de investigación) adjuntan una columna con un valor de p para evaluar la significancia estadística de eventuales diferencias que pudieran hallarse. Muchos autores consideran que este valor de p no posee valor alguno y que si la aleatorización y asignación de grupos fue adecuada, entonces cualquier diferencia entre los grupos será debida al azar, inclusive aún sin que el test de hipótesis arroje una $p < 0,05$.⁽²⁾ El abordaje ideal al hallar una diferencia en el balance de factores pronósticos entre los grupos en estudio (control y tratamiento) es determinar si al analizar los resultados del ensayo se realizó ajuste por aquellos potenciales factores pronósticos importantes. Si los análisis ajustados y sin ajustar arrojan resultados similares, podemos estar tranquilos que las diferencias observadas en la tabla de características basales no son relevantes ni modifican el efecto de la intervención sobre el evento de interés.

¿Se mantuvo el balance de los factores pronósticos a lo largo del estudio?

1) *¿El grupo al que fue asignado cada sujeto se mantuvo oculto (ciego)? ¿es un estudio abierto? ¿Quiénes fueron cegados: el paciente, el investigador y personal del estudio, el evaluador de puntos finales, todos?*

Una estrategia para sostener el equilibrio o balance entre los factores pronósticos de los grupos durante el transcurso del estudio es mantener oculto la rama o grupo al que cada sujeto fue asignado. Los estudios que poseen este diseño se denominan “ciegos”, y se puede cegar tanto a los pacientes, los investigadores y a los evaluadores de puntos finales, como así también a todos a la vez. De esta manera, se puede evitar que el reporte de síntomas por parte de los pacientes o la adherencia al tratamiento sean influenciados por el conocimiento sobre el régimen al que fue aleatorizado. De la misma forma, los diseños “ciegos” evitan que el investigador reporte o interprete más eventos en un grupo que en el otro influenciado por sus creencias sobre la eficacia de la intervención en estudio.

Cuanto más relevante será el juicio clínico para determinar la ocurrencia de un evento, más importante será que el estudio posea un diseño ciego. Así, cuando el punto final de un estudio es “poco influen-

ciable” por interpretaciones o juicios individuales (ejemplo: mortalidad por todas las causas) menos relevante será el uso de un diseño ciego.

En algunos casos en que tanto los pacientes como los investigadores no puedan ser cegados (por ejemplo en algunos estudios quirúrgicos), lo recomendable es que quien adjudique la ocurrencia de los puntos finales del estudio sea una tercera persona, que si se encuentre ciega.

Es importante además de analizar si el estudio posee un diseño ciego, si los grupos que se están comparando en el ensayo recibieron igual tratamiento aparte de la intervención en estudio. En muchas oportunidades nos encontraremos que el estudio posee un diseño “add on”, en el que el tratamiento experimental se agrega a una serie de tratamientos que ya han demostrado su eficacia. Por ejemplo, si quisiéramos evaluar el efecto adicional de una nueva droga para evitar mortalidad en pacientes con insuficiencia cardíaca, ambos grupos deberían recibir tratamiento con beta-bloqueantes, enalapril, aspirina, furosemida y espironolactona. Los participantes del estudio deben tener igual oportunidad de acceso a las co-intervenciones de las que pueden obtener algún beneficio independientemente del grupo al que fueron aleatorizados.

2) *¿Fueron analizados todos los sujetos participantes dentro de los grupos a los que fueron originalmente aleatorizados? (principio de intención de tratar)*

Porque cualquier situación que pudiera ocurrir luego de la aleatorización puede alterar la chance que un participante del estudio desarrolle un punto final de interés, es importante que todos los pacientes (aún aquellos que no recibieron el tratamiento al que fueron aleatorizados, o que tanto accidental como voluntariamente recibieron el tratamiento inadecuado) sean analizados de acuerdo al grupo que fueron aleatorizados. Esto se denomina “principio por intención de tratar”, y sostiene que una vez que obtenemos dos grupos comparables al inicio del ensayo, debemos mantenerlos para preservar el poder de la aleatorización. En esta sección deberemos ser cautos y fijarnos que no solo esté descripta en la sección de métodos la forma en que se analizó el punto final primario, si no también que así se haya realizado verdaderamente.

Los participantes de un estudio que han abandonado

el mismo, o que por algún motivo no fueron “adherentes” al tratamiento asignado, pueden poseer diferente pronóstico del resto de los sujetos que permanecieron en el estudio. Y más aún, las causas de su no-adherencia pueden estar vinculadas al tratamiento asignado o procedimientos del estudio. Incluso, es reconocido que los pacientes no-adherentes al tratamiento asignado poseen generalmente diferente evolución que los adherentes, aun siendo pacientes asignados al grupo “placebo”.

Cuando los investigadores deciden analizar los resultados de un estudio de acuerdo al tratamiento que realmente recibió cada sujeto participante, el análisis que se realiza se denomina “per protocol” o de acuerdo a protocolo. Esto es útil cuando lo que se evalúa son los efectos adversos potenciales de un tratamiento.

¿Los grupos se mantuvieron equilibrados al finalizar el estudio?

1) *¿El seguimiento de los pacientes fue suficientemente largo y completo? ¿y similar en ambos grupos?*

Una vez que estemos satisfechos con que el estudio fue aleatorizado deberemos revisar si todos los pacientes que ingresaron al mismo fueron contemplados al analizar los resultados. Idealmente, al final de un estudio de investigación, se debiera conocer el status de cada uno de los sujetos participantes en relación al punto final de interés primario. Cuantos más pacientes haya de los cuales se desconozca que ha ocurrido con ellos al finalizar el estudio (pérdida de seguimiento), más se verá amenazada la validez del estudio. Esto se debe a que los pacientes que se han perdido al seguimiento pueden poseer diferente pronóstico que los que han permanecido en el estudio.

No existe un valor a partir del cual podemos considerar que ha habido una importante pérdida de seguimiento que amenace la validez del estudio. Algunos autores sugieren que hasta un 20% de pérdida al seguimiento podría ser aceptable. Pero para hacer una adecuada valoración de este valor debemos tener en cuenta el número total de sujetos incluidos en el estudio y el número de eventos totales. Una manera de estimar el posible impacto de la pérdida de seguimiento en los resultados de un ECC es asumir que ha ocurrido el “peor escenario”. En este ejercicio asumimos que todos los pacientes que se

han perdido en el grupo tratamiento han padecido el evento de interés, mientras que a su vez todas las pérdidas de seguimiento del grupo control no han sufrido ningún evento, y rehacemos el cálculo de los resultados del estudio. Si no existen modificaciones relevantes en los resultados podemos asumir que la pérdida de seguimiento no fue importante en este estudio.

Otro aspecto importante a considerar es que la longitud del seguimiento planeada por el estudio haya sido lo suficientemente larga como para permitir ver los eventos o puntos finales de interés.

Esto es importante también para la consideración de los riesgos potenciales de los tratamientos experimentales cuando los llevamos a la práctica diaria. Por ejemplo, ningún estudio realizado con inhibidores selectivos de la recaptación de serotonina ha tenido una ventana de seguimiento mayor a 1 año, sin embargo, muchas veces estas drogas suelen utilizarse por períodos más largos.

2) ¿Se detuvo el estudio antes de tiempo?

A pesar que cada vez es más frecuente observar esto en la literatura médica, la detención temprano de estudios de investigación cuando se observa un aparente beneficio puede ser riesgo para la interpretación adecuada de los resultados.¹² Generalmente, la detención temprana de estudios genera una magnificación del beneficio observado. Esto puede ser especialmente problemático cuando el tamaño del estudio o el número de eventos es pequeño.

Conclusiones

En esta entrega hemos repasado los aspectos más importantes al analizar la validez de ECC. El proceso sistemático y metódico de análisis nos mostrará cuan confiables pueden ser las conclusiones de un estudio.

La aplicabilidad de los resultados de un ensayo a la práctica clínica requiere un importante conocimiento de nuestros pacientes cotidianos, incorporando a la toma de decisiones los valores y preferencias de estos.

Bibliografía

1 Bottaro FJ. Diseño de los estudios de investigación. Debilidades y fortalezas. *Hematología*. 2014; 18 (1): 74-83.

2 Straus SE, Galsziou P, Richardson SW, Haynes BR. Evidence-based Medicine. How to practice and teach it. Churchill Livingstone Elsevier; 2011. Chapter 4: Therapy. p.67-136.

3 Guyatt G, Straus SE, Meade M, Kunz R, Cook DJ, Devereaux PJ, Ioannidis J. Therapy (Randomized Trials). In: Guyatt G, Drummond R, Meade MO, Cook DJ, editors. *User's Guide To The Medical Literature. A Manual for Evidence-Based Clinical Practice*. JAMA & Archives Journals, American Medical Association; 2008. p. 67-86.

4 Bottaro FJ. ¿Cómo leer la literatura científica? *Hematología*. 2013; 17 (1): 79-85.

5 Stampfer MJ, Colditz GA, Willett WC, y col. Postmenopausal estrogen therapy and cardiovascular disease.

Ten-year follow-up from the nurses' health study. *N Engl J Med*. 1991; 325(11):756-62.

6 Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med*. 1991; 20(1):47-63.

7 Hulley S, Grady D, Bush T, y col. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA*. 1998 Aug 19;280(7):605-13.

8 Rossouw JE, Anderson GL, Prentice RL, y col; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*. 2002; 288(3):321-33.

9 Bjelakovic G, Nikolova D, Simonetti RG, y col. Antioxidant supplements for prevention of gastrointestinal cancers: a systematic review and meta-analysis. *Lancet*. 2004; 364(9441):1219-28.

10 Echt DS, Liebson PR, Mitchell LB, y col. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991; 324(12):781-8.

11 Packer M, Carver JR, Rodeheffer RJ, y col. Effect of oral milrinone on mortality in severe chronic heart failure. The PROMISE Study Research Group. *N Engl J Med*. 1991; 325(21):1468-75.

12 Montori VM, Devereaux PJ, Adhikari NK, y col. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203-9.