

Trabajos sobre tratamientos: cómo analizar su validez (y cuándo confiar en sus resultados)

ANA GABRIELA PALIS

Para analizar críticamente los resultados de un estudio sobre tratamiento nos concentramos fundamentalmente en tres aspectos¹:

- 1) Su validez
- 2) Su importancia y precisión
- 3) Su aplicabilidad al cuidado de nuestro paciente particular

En este artículo veremos cómo evaluar la **validez** de los resultados de un estudio sobre tratamiento.

¿Qué significa que los resultados de un trabajo son “válidos”?

La validez de un estudio es la cualidad por la cual éste representa una estimación imparcial de la verdad subyacente². Es decir que *lo que el trabajo estima para la población estudiada se aproxima a la verdad en la población general*.

Este principio apunta a la credibilidad en los resultados de un estudio: intentamos discernir si representan una estimación sin sesgos del efecto del tratamiento o si han sido influidos de algún modo para llevar a una conclusión poco confiable de ser verdadera.

Si decidimos que los resultados son válidos y el estudio muestra una evaluación no sesgada del efecto del tratamiento, entonces justificará continuar examinando los resultados³.

¿Qué es el “sesgo”?

Sesgo (*bias* en inglés) es aquello que hace que el resultado de una observación se aparte sistemáticamente de la verdad. Los sesgos en la investigación pueden tener diferentes orígenes. Uno de los más importantes —el “sesgo de selección”— ocurre cuando se realizan comparaciones entre grupos que difieren en cuanto a sus “factores pronósticos” (y no en el tratamiento que se está estudiando)².

¿Qué es un “factor pronóstico”?

Es una característica que aumenta o disminuye el riesgo de un evento (positivo o adverso). Por ejemplo: la edad en el glaucoma (la edad avanzada aumenta el riesgo de glaucoma), el sexo en algunos tipos de desprendimiento de retina (los hombres jóvenes miopes tienen riesgo aumentado de desprendimiento de retina regmatógeno), la hipertensión en la oclusión de vena central de la retina (el no ser hipertenso disminuye el riesgo), entre otros.

La distribución equilibrada de los factores pronósticos, tanto conocidos como no conocidos, en ambos grupos (de tratamiento y control) convierte a éstos en similares salvo por la intervención. Esta similitud en la distribución de los factores pronósticos nos permitirá comparar los resultados de las intervenciones en ambos grupos ya que estos resultados, a favor o en contra del tratamiento, se relacionarán con éste y no con características disímiles que puedan tener los grupos entre sí.

¿Dónde buscamos la “validez” de un estudio?

Que los resultados de un estudio sean válidos depende del hecho de que éste haya sido diseñado y realizado de manera que podamos confiar en lo que afirma sobre los beneficios o desventajas de un tratamiento¹. Este diseño se evalúa en la sección de **métodos** del estudio.

¿Cómo evaluamos la validez de un estudio?

Podemos agrupar la evaluación de los métodos de un estudio en ocho preguntas¹ divididas en dos grupos (tabla):

1. El primer grupo de preguntas nos ayuda a decidir si las personas expuestas al tratamiento experimental tenían un pronóstico similar a los

Tabla 1. Planilla de evaluación de la validez de trabajos sobre tratamiento.

Características	Pregunta	Sí	No	No dice
Los grupos control y experimental, ¿comenzaron el ensayo con pronóstico similar?	¿Los pacientes fueron aleatorizados?			
	¿La asignación fue oculta (ciega o enmascarada)?			
	¿Se asignaron los pacientes en los que fueron aleatorios?			
	¿Los pacientes en los grupos de tratamiento y control eran similares en cuanto a factores pronósticos conocidos?			
Los grupos control y experimental, ¿mantuvieron un pronóstico similar después de comenzado el estudio?	¿Los pacientes desconocían la asignación al grupo?			
	¿Los médicos desconocían la asignación al grupo?			
	¿Los evaluadores de resultados desconocían la asignación al grupo?			
	¿Se completó el seguimiento?			

pacientes expuestos a una intervención control **al comenzar** el estudio.

2. El segundo grupo nos ayuda a confirmar que los dos grupos permanecieron **similares** con respecto de los factores pronósticos **a lo largo** de todo el estudio.

Cuantas más respuestas positivas obtengamos, mayor será la validez del estudio y, por lo tanto, más confianza podremos tener en que sus resultados tienen poco o ningún sesgo.

Primer grupo de preguntas

Son las que nos ayudan a decidir si las personas expuestas al tratamiento experimental tenían un pronóstico similar a los pacientes expuestos a una intervención control al comenzar el estudio.

Pregunta 1: ¿Los pacientes fueron aleatorizados?

Lo primero que debemos revisar es si los pacientes fueron aleatorizados (*randomized*, en inglés), es decir, si fueron ubicados en los grupos de tratamiento y control mediante algún proceso por el cual cada paciente tuvo la misma oportunidad de aparecer en el grupo tratamiento o en el control (semejante pero no igual al efecto de distribuirlos al azar)⁴. Este proceso equilibra la distribución de los factores pronósticos en ambos grupos.

Si, por el contrario, los factores pronósticos se encuentran desequilibrados entre los grupos de tratamiento y control, los resultados del estudio estarán sesgados y sub- o sobreestimarán el efecto del tratamiento.

Por ejemplo: un grupo de investigadores concluye que el tratamiento con una nueva droga mejoró la progresión

de una membrana neovascular coroidea en una población de pacientes comparado con el tratamiento con una droga ya en uso (que se aplicó al grupo control). Si el trabajo no especifica que los pacientes fueron aleatorizados, el lector no podrá confiar en que estos resultados representen la verdad en la población general, ya que quizás los pacientes que se incluyeron en el grupo control tenían una maculopatía más avanzada o eran fumadores o de mayor edad. Este grupo presentaría peor pronóstico desde el momento previo a la intervención en estudio y no puede confiarse en que el mayor o menor efecto de la droga sea solamente debido al tratamiento o influyó este desequilibrio en el pronóstico de ambos grupos.

El proceso de aleatorización reduce el peso de estos factores pronósticos. Si el azar asigna a un participante del estudio a un brazo del mismo todos los factores estarán equitativa y equilibradamente distribuidos en ambos grupos. Estos incluyen factores *conocidos*, tales como la edad, el hábito tabáquico, el estadio de la maculopatía, pero sobre todo factores *desconocidos* que puedan afectar los resultados. De esta manera, la randomización permite que los grupos sean similares **salvo por la intervención**⁶.

Resumiendo, la aleatorización genera:

- Equilibrio tanto en factores pronósticos conocidos como desconocidos, de tal forma que la única diferencia entre los grupos en estudio sea la intervención.

- Grupos comparables entre sí.

Es debido a esto que los ensayos controlados aleatorizados son los diseños de estudios sobre tratamientos que, en general, menor sesgo poseen (tienen mayor validez intrínseca)⁷.

Pregunta 2: ¿Se ocultó la asignación de los pacientes a los grupos durante la aleatorización?

Si los investigadores conocen el brazo del estudio al cual cada paciente fue asignado (es decir, si la aleatorización no es oculta) podrían reclutar pacientes más —o menos— enfermos a cada grupo de tratamiento o control, alterando la aleatorización y el estudio arrojaría un resultado sesgado. El proceso de aleatorización puede ocultarse, por ejemplo, a través de la preparación de medicación en frascos sin nombre en un laboratorio, la aleatorización remota (quien recluta al paciente llama a un centro de métodos para encontrar el brazo del estudio al cual fue asignado) o mediante sobres sellados.

Pregunta 3: ¿Fueron los pacientes analizados en el grupo al cual estaban primariamente aleatorizados?

La aleatorización puede también corromperse si se omiten de los resultados a pacientes que no utilizan, por motivos diversos, el tratamiento que les fue asignado. Debido a que las razones por las cuales las personas que no toman su medicación se relacionan frecuentemente al pronóstico. Excluir del análisis a pacientes no cumplidores deja afuera a quienes podrían estar destinados a tener un mejor resultado y destruye la comparación no sesgada que provee la aleatorización. Este principio para preservar el valor de la aleatorización de atribuir a todos los pacientes al grupo al cual fueron asignados, en lugar de al del tratamiento que realmente recibieron, se denomina **análisis por intención de tratar**⁵.

Por lo tanto, al revisar el informe de un ensayo aleatorizado debe buscarse evidencia de que los investigadores analizaron a todos los pacientes en los grupos a los cuales fueron originalmente asignados.

Pregunta 4: ¿Eran similares en cuanto a factores pronósticos los pacientes del grupo tratamiento y los del grupo control?

Debido al azar, a veces la aleatorización falla en alcanzar el propósito de crear grupos cuyo pronóstico, con respecto del resultado de interés, sea similar, siendo esto aún más probable para muestras pequeñas.

Al leer un trabajo deben mirarse las características de los pacientes con tratamiento o control al comenzar el estudio. Aunque el lector nunca sabrá si existe similitud para los factores pronósticos desconocidos podrá confiar en que los factores pronósticos conocidos están equilibrados⁵.

Segundo grupo de preguntas

Son las que nos ayudan a decidir si los grupos de tratamiento y control se mantuvieron similares con respecto de los factores pronósticos a través del estudio.

Si la aleatorización tiene éxito, los grupos de tratamiento y control comienzan el estudio con un pronóstico muy similar. Sin embargo, esto no garantiza que los dos grupos permanezcan equilibrados en cuanto al pronóstico a lo largo de todo el ensayo, ya que diferencias en el proceso de cuidado de los pacientes en aspectos que no sean la intervención en estudio pueden sesgar los resultados.

Pregunta 5: ¿Desconocían los pacientes el grupo al cual habían sido asignados?

Los pacientes que realizan un tratamiento que creen es eficaz pueden sentir y evolucionar mejor que aquellos que no lo creen, aún si el tratamiento no tiene acción biológica (efecto placebo). También pueden contestar preguntas o realizar exámenes de manera diferente, dependiendo de si creen que están tomando medicación activa o no⁵.

La mejor manera de evitar estos problemas es asegurarse de que los pacientes desconozcan si están recibiendo o no el tratamiento experimental. Por ejemplo, los pacientes del grupo control pueden recibir un comprimido o un frasco de gotas inertes que sea idénticos en color, sabor y consistencia a la medicación activa administrada al grupo de tratamiento.

Pregunta 6: ¿Desconocían los investigadores el grupo al cual habían sido asignados los pacientes?

Puede tenerse más confianza en los resultados de un estudio si los investigadores estaban “a ciegas” sobre qué pacientes recibieron tratamiento activo o son parte del grupo control. El enmascaramiento elimina la posibilidad de que se administren cointervenciones diferentes a ambos grupos, ya sea en forma consciente o inconsciente.

Pregunta 7: ¿Desconocían los evaluados el grupo al cual habían sido asignados los pacientes?

Si los integrantes del equipo de investigación que mide o registra resultados no están “a ciegas” pueden proveer interpretaciones diferentes de sus hallazgos o alentar de manera diferente a los pacientes durante evaluaciones de desempeño o calidad de vida, todo lo cual puede distorsionar los resultados⁵.

Casi siempre puede mantenerse a ciegas al personal evaluador del estudio, aún en el caso de no poder hacerlo con los pacientes y médicos tratantes (como en el caso de tratamientos quirúrgicos).

Pregunta 8: ¿Se completó el seguimiento?

Los pacientes cuyo resultado es desconocido al finalizar el estudio se denominan “pérdidas de seguimiento”. Cuanto mayor es su número, más se compromete potencialmente la validez de un estudio, ya que los pacientes que se

pierden suelen tener diferente pronóstico de aquellos que permanecen. Estos pacientes pueden desaparecer porque sufren efectos adversos (inclusive la muerte) o porque evolucionan bien (y entonces no vuelven para ser evaluados).

Se considera que la validez de un estudio suele comprometerse cuando la pérdida de seguimiento es mayor del 20% del total de la población que comenzó el estudio. En estos casos pueden analizarse los resultados de acuerdo con lo que se denomina "escenario del peor caso" o "peor escenario", que consiste en asumir que los pacientes perdidos sufrieron el evento adverso de interés. Si el peor escenario altera significativamente los resultados, la validez está comprometida⁵.

Conclusiones

La evaluación de la validez de un estudio sobre tratamiento, luego de responder a las preguntas enumeradas, permite decidir si justifica continuar analizando sus resultados. Ayuda también a evaluar y ponderar estos resultados y conclusiones a la luz de la rigurosidad científica de su diseño metodológico. La decisión final sobre la validez de un estudio raramente es un "sí" o "no" neto, sino un rango entre estudios fuertes muy probables de ofrecer una estimación precisa del efecto del tratamiento, a estudios

débiles que conducen a una estimación sesgada del efecto. La decisión de en qué parte de este rango se encuentra un estudio implica cierta subjetividad para decidir en qué grado nos ayudará a recomendar un tratamiento determinado a un paciente⁵. Las respuestas a las preguntas presentadas en la tabla facilitan y agilizan la toma de estas decisiones.

Bibliografía

1. Guyatt G, Cook D, Devereaux PJ, Meade M, Straus S. Therapy. En: Guyatt G, Rennie D. (eds.). *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago:AMA Press, 2002, p. 58-66.
2. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials*. 2nd ed. Baltimore:Williams & Wilkins, 1988, p. 7-8.
3. Sackett D, Straus S, Richardson W. *Evidence-based medicine: how to practice and teach EBM?* 2nd ed. Edinburgh: Churchill Livingstone, 2000. p. 105-110.
4. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials*. 2nd ed. Baltimore:Williams & Wilkins, 1988, p. 143.
5. Guyatt G, Cook D, Devereaux PJ, Meade M, Straus S. Therapy. In: Guyatt G, Rennie D, editors. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago:AMA Press, 2002, p. 267-73.
6. Alderson P, Green S (eds.). Cochrane Collaboration open learning material for reviewers. Version 1.1 [November 2002]; Module 1. p 4-5. http://www.cochrane-net.org/openlearning/PDF/Module_1.pdf [consultado: 6 octubre 2006].
7. Higgins JPT, Green S (eds.). *Cochrane handbook for systematic reviews of interventions* 4.2.6. [updated September 2006]; Section 4. p. 60, 79-81. <http://www.cochrane.org/resources/handbook/hbook.htm> [consultado: 6 octubre 2006].