

EL ENSAYO CLÍNICO COMO MÉTODO DE INVESTIGACIÓN

VERÓNICA AHUMADA P.¹ y DAVID TORRES P.²

- La principal ventaja de un ensayo clínico es su capacidad de demostrar causalidad: la aleatorización disminuye la probabilidad de variables confundentes y el ciego disminuye el riesgo que los efectos se deban a una co-intervención o a una evaluación sesgada.
- Los ensayos clínicos son útiles en la medida en que nos permiten extraer inferencias válidas no sólo respecto a lo que pasó en los pacientes del estudio, sino a la generalización de estos resultados a los pacientes que están fuera del estudio.
- Un ensayo clínico debe ser previamente diseñado en forma minuciosa, con una metodología adecuada teniendo en cuenta una serie de variables para minimizar todo sesgo, desde la selección de pacientes hasta la expresión de los resultados.
- Los sesgos más frecuentes son los de selección y de medición.
- Los estudios mal diseñados hacen que los resultados parezcan mejores de lo que realmente son o que tratamientos inadecuados parezcan funcionar, cuando en realidad no lo hacen o incluso pueden ser perjudiciales.

Probablemente el primer ensayo clínico que se conoce fue realizado por el médico escocés, James Lind a bordo del HMS Salisbury de la armada británica. Usó a doce marineros que padecían de escorbuto y los dividió en seis parejas, de manera que cada pareja recibiera un tratamiento determinado diferente al de los demás manteniendo una dieta basal similar en todos; de esta manera concluyó que dar limones y naranjas disminuía los síntomas de la enfermedad, pues los marineros que las habían recibido estuvieron listos para trabajar al cabo de pocos días. No llegó a la conclusión, como muchas

veces erróneamente se señala, que la causa del escorbuto era una carencia alimentaria.

En la actualidad, escuchar hablar sobre ensayos clínicos es algo cotidiano para cualquier médico. Múltiples tratamientos o procedimientos que se realizan a diario están basados o surgieron debido a ellos y existe la impresión que es el estándar de oro para responder cualquier pregunta. Pero, ¿Tenemos claro realmente qué es un ensayo clínico, cómo se hace?, ¿Es conveniente hacer un ensayo clínico para responder a una determinada pregunta?, ¿Cómo evaluar si el ensayo clínico que tenemos al frente es de calidad?. El objetivo de este artículo es orientar a los médicos en la lectura crítica de este tipo de investigación para responder a estas preguntas, haciendo énfasis en la metodología y potenciales sesgos que se encuentran frecuentemente en la literatura médica.

El ensayo clínico es un estudio experimental prospectivo, que compara un tratamiento (intervención) con un grupo control (placebo o activo), y luego observa su efecto o resultado (Figura 1)¹. La principal ventaja es su capacidad de demostrar causalidad. La asignación aleatoria de la intervención disminuye la probabilidad de variables confundentes (al hacer los grupos comparables) y el ciego disminuye el riesgo que los efectos se deban a una co-intervención o a una evaluación sesgada. Para leer críticamente un ensayo clínico se debe analizar la validez interna (o el diseño del estudio), los resultados y su aplicabilidad o validez externa.

Los ensayos clínicos en general son caros, consumen mucho tiempo, la pregunta que pretenden responder es relativamente específica y potencialmente exponen a los pacientes a daño. El estudio clínico ha sido considerado por muchos años el patrón de oro o *gold standard* para establecer la efectividad de las intervenciones; actualmente disputa este lugar en la cúspide de la evidencia con el metaanálisis obtenido a partir de estudios pros-

¹ Interna Medicina, Universidad de los Andes.

² Clínica Santa María, Instructor, Universidad de los Andes.

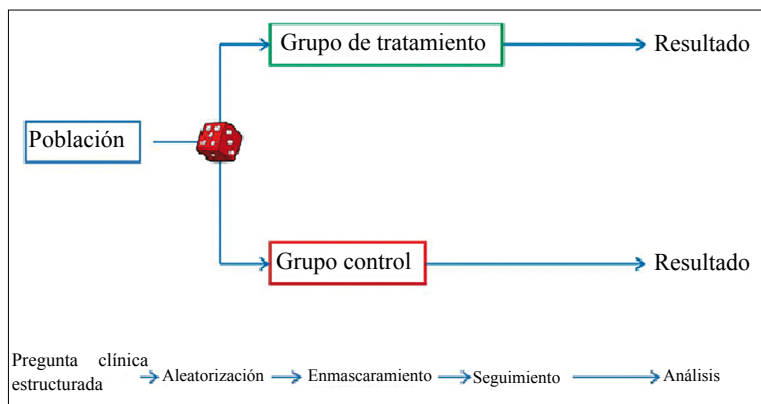


Figura 1. Diagrama de flujo de un ensayo clínico.

pectivos. Sin embargo, hay muchas preguntas que no pueden ser respondidas por este tipo de diseño, donde los estudios observacionales son una mejor elección (por ejemplo, estudios de hipertermia maligna) y muchos ensayos clínicos son de tamaño reducido, tienen problemas de diseño que limitan sus conclusiones y que impiden establecer si una intervención es efectiva. Esto hace necesario esperar que otros ensayos clínicos obtengan resultados similares; de aquí la utilidad de revisiones sistemáticas de ensayos clínicos o bien que los ensayos clínicos se hagan con una gran cantidad de pacientes, que permitan establecer relaciones menos tentativas entre la intervención realizada y el resultado obtenido.

El análisis de validez interna incluye los elementos tangibles del plan de estudio: la pregunta de investigación, el diseño apropiado, las características de los pacientes y de las mediciones, cálculo de tamaño muestral, etc. El objetivo del investigador es disponer de estos elementos de manera de hacer que el proyecto funcione eficientemente (rápido, fácil y barato)². Por otro lado, los estudios son útiles en la medida en que nos permiten extraer inferencias válidas. Estas inferencias son útiles no sólo respecto a lo que pasó en los pacientes del estudio, sino a la generalización de estos resultados a los pacientes que están fuera del estudio y que tienen características similares. El objetivo es minimizar los errores, aleatorios y sistemáticos, que puedan “amenazar” las conclusiones en que se basaron estas inferencias.

DISEÑO

El diseño de los ensayos clínicos, así como de cualquier trabajo de investigación, requiere una planificación previa y una protocolización estricta.

Para fines didácticos, las nombraremos en el orden en que van apareciendo a medida que uno lee un artículo, aunque este orden no corresponda necesariamente en la secuencia temporal.

La pregunta de investigación es el objetivo del estudio, lo que el investigador quiere conocer y, generalmente, parten como un concepto general que después se transforma en algo más concreto. Esta pregunta no sólo debe estar bien estructurada (en la cual se incluya pacientes, la intervención, la comparación y el resultado), sino que también debe ser factible, interesante, novedosa, ética y relevante. Ejemplo: “El fármaco A (intervención) reduce la incidencia de náuseas y vómitos postoperatorios (resultado esperado) al ser administrado en forma profiláctica 4 horas antes del procedimiento en pacientes sometidos a colecistectomía electiva (pacientes), cuando se compara con el fármaco B (comparación)”.

Luego se debe revisar la selección de los participantes: lo primero, especificar el criterio de selección que define el tipo de paciente más apropiado para responder la pregunta y, lo segundo, determinar cómo van a ser reclutados. Para lo primero, hay que establecer criterios de inclusión y exclusión que debieran mejorar la probabilidad de encontrar el resultado primario del estudio. Si el resultado es un evento raro, como la enfermedad tromboembólica perioperatoria, se hace necesario seleccionar pacientes que tienen una alta probabilidad de presentar el resultado de estudio. Mientras más estrictos los criterios de selección más se reduce la generalización de los resultados (validez externa) y hace que el reclutamiento sea más difícil. Para lo segundo, hay que definir el número de pacientes necesarios para responder la pregunta de investigación y finalmente definir cómo y dónde se va a realizar el reclutamiento de los pacientes, para lograr una muestra representativa.

Dentro de toda la población definida como elegible de acuerdo a los criterios establecidos, se debe asignar aleatoriamente (al azar) los grupos: el que recibirá la intervención y el grupo control con el cual se compararán los resultados. De esta manera se consigue que cada participante tenga la misma probabilidad de estar en el grupo 'intervención' o en el grupo 'comparación'. La inclusión al estudio y la asignación a los grupos son fuentes frecuentes de error sistemático o sesgo y se deben realizar todos los esfuerzos para evitarlo y, el lector crítico, todos para detectarlos. Esto debe hacerse usando una secuencia oculta (enmascarada) de aleatorización, que elimina el **sesgo de selección** en los ensayos clínicos. Si esta es bien realizada (por computador, o incluso tirando una moneda al aire), se evita cualquier intervención consciente o inconsciente del investigador. La aleatorización inadecuada puede exagerar tremendamente la estimación del efecto; estudios que comparan aleatorización adecuada con la no descrita o inadecuada, pueden mostrar diferencias del 30 - 40% de la magnitud del efecto^{3,4}.

Elegir las variables que se van a evaluar y comparar: generalmente se eligen 4 o más variables epidemiológicas para comparar los grupos en forma basal, además de las variables de resultado. En estas variables se deben incluir todas aquellas que se sabe inciden en los resultados. Las variables de resultado deben tener relación con los objetivos del estudio y ser relevantes para los pacientes; no sirve sólo medir los resultados que reflejan la efectividad de la intervención, sino también los de seguridad (incidencia de efectos adversos). Por ejemplo, si estudiamos un relajante muscular nuevo, la latencia y la duración del efecto son resultados de efectividad relevantes, pero la incidencia de hipotensión, reacciones anafilácticas o mortalidad por cualquier causa (resultado de seguridad) puede ser incluso más relevante, dependiendo de su magnitud.

Para minimizar el **sesgo de medición** también debe usarse el enmascaramiento o ciego. Este consiste en una serie de medidas o precauciones que se toman con el fin de que, a lo largo de todo el estudio, el paciente, el médico o ambos, además de aquellos que evalúan la respuesta a los tratamientos, desconozcan a qué grupo pertenece cada paciente⁵. El objetivo del ciego es evitar la co-intervención que puede darse en dos situaciones: la primera, cuando el investigador conoce que tratamiento recibe cada paciente y presta especial atención al grupo experimental en el cual interviene con el afán de encontrar elementos que validen su hipótesis; la segunda, cuando los pacientes del grupo control saben que no están recibiendo más que placebo y por su

cuenta toman otras medidas que pueden afectar el resultado. Adicionalmente, con el ciego es posible limitar el sesgo del evaluador, el que al no saber a qué grupo pertenece el paciente, no puede sobreestimar el efecto de la intervención (o del control): los estudios no ciegos sobreestiman los efectos del tratamiento en un 17%⁶ e incluso pueden cambiar totalmente las conclusiones³.

Debido a la gran magnitud del sesgo esperado en los estudios que no son aleatorizados ni ciegos, se ha utilizado puntajes que reportan la calidad metodológica de los estudios, con hincapié en estos dos puntos; los estudios con menor calidad, consistentemente sobreestiman los efectos del tratamiento⁷.

Definir la intervención: hay que considerar efectividad y seguridad, definir si se harán una o varias intervenciones, posibilidad de ciego y protocolizar el tratamiento para llevarlo a cabo en la práctica clínica. Considerando lo anterior, lo mejor es que haya una única intervención para evitar la duda sobre a cual de todas las intervenciones atribuir el resultado obtenido, que la intervención use la menor dosis efectiva posible para así evitar efectos adversos, que sea posible el ciego y que se pueda protocolizar la terapia para el uso en la práctica clínica.

Grupo control: lo ideal es que los pacientes que lo constituyen no reciban ningún tratamiento activo, pero sí placebo idéntico al del grupo intervenido. Lo descrito no siempre es posible. Por ejemplo, en un estudio en el que se quiere evaluar la utilidad de las estatinas en la prevención de infarto agudo del miocardio en pacientes con antecedentes de enfermedad coronaria; a los participantes del grupo control puede no dárseles estatina, pero sería poco ético negarles el consumo de beta-bloqueadores, aspirina y aquellos fármacos de los que se beneficiarán aunque su uso vaya a influir en los resultados, por lo que si los pacientes usan otras terapias es preciso aumentar el tamaño de la muestra, hacer un seguimiento más largo, y considerar ajustes estadísticos⁸. En el caso que exista una terapia estandarizada para la enfermedad a tratar, el nuevo tratamiento debe compararse con aquello ya probadamente eficaz, ya que es éticamente incompatible dejar a un grupo de pacientes sin un tratamiento comprobadamente efectivo. Sin embargo, siguen apareciendo estudios que comparan las nuevas intervenciones con placebo.

Los estudios clínicos deben tener un cálculo de tamaño muestral que permite estimar la cantidad de pacientes (n) necesarios para lograr demostrar como significativas las diferencias esperadas entre los dos grupos. Si el valor de la diferencia es alta, el

n puede ser pequeño. Es mucho más probable que la diferencia sea pequeña, por lo que el n debe ser mucho mayor para lograr mostrar una diferencia. Pese a esto, los ensayos clínicos son absurdamente pequeños. En general, los estudios con menos de 10-15 pacientes no debieran considerarse.

El azar también puede ser un factor relevante, pese al cálculo de tamaño muestral adecuado. Estudios con poblaciones e intervenciones similares pueden tener resultados muy distintos. El grado de variabilidad entre los estudios es muy alto, ya que el cálculo de tamaño muestral se realiza para detectar una determinada diferencia, pero la magnitud de esta diferencia estimada como importante puede ser muy diferente⁹.

Una vez hecho lo anterior, el investigador debe encontrar pacientes que puedan lograr adherirse al protocolo para que sea bien llevado por los participantes, ya que el efecto esperado de la intervención se reduce en la medida en que los participantes no reciben el tratamiento correspondiente al grupo asignado¹⁰. Para ello existen múltiples estrategias tales como etiquetar las pastillas, llamar por teléfono, hacer visitas, etc. Siempre tener en cuenta que se logra mayor adherencia cuando los exámenes de control son indoloros y no invasivos y percibidos por los pacientes como beneficiosos. Aquellos pacientes que no siguen el protocolo a pesar de las medidas tomadas, igual deben ser seguidos y considerados en los resultados, lo que se denomina Intención de tratar. El objetivo es analizar la intervención de una manera lo más aproximada a la realidad y permite detectar algunas razones por las que la adherencia es incompleta (incomodidad, efectos adversos, etc.).

Pese a los esfuerzos de editores y revisores, algunos artículos son simplemente incorrectos. Esto se puede deber a muchas razones, pero principalmente dos: incorrección estadística y manipulación de datos. La incorrección estadística puede disfrazarse de muchas formas: lo más sencillo de detectar es que un resultado presentado como estadísticamente significativo no lo es, o detectar que se usó el test estadístico inadecuado para la variable estudiada. En otras ocasiones los autores hacen una “pesca de arrastre” con los datos, donde buscan una significación estadística y escriben el artículo en función de ella. La recomendación es leer los resultados con cautela, especialmente cuando los estudios son pequeños o hay muy poca diferencia en los resultados de cada grupo.

La parte escrita también puede llevar a error. Aunque las pruebas estadísticas no muestren diferencias, no es infrecuente ver los resultados presentados como un éxito de la intervención, lo que

puede llevar al lector ocasional a engañarse con los resultados, especialmente cuando se lee sólo el resumen de los artículos. Esta práctica sólo debe limitarse a decidir si se sigue leyendo el artículo o no; basar opiniones o conductas sólo en la lectura de un resumen es una práctica condenable.

La manipulación de resultados puede ser más compleja de detectar. En general los pacientes no se describen adecuadamente y en algún momento se aprecia una diferencia significativa por algún test poco utilizado. El objetivo es hacer “positivo” lo “negativo” y la dirección del sesgo favorece la nueva intervención. Debe siempre sospecharse cuando hay muy poca diferencia entre los resultados.

RESULTADOS

Los resultados reportados pueden ser problemáticos. Frecuentemente se utilizan resultados indirectos o secundarios presentados como clínicamente relevantes, lo que da una falsa sensación de efectividad. Deben considerarse todos los resultados relevantes, no sólo el que tiene relación directa con lo que se está investigando, sino también efectos adversos y factores que demuestran un beneficio real del tratamiento en cuestión. Por ejemplo, un estudio mostró que un nuevo hipoglucemiante efectivamente lograba un mejor control glicémico, pero los pacientes tenían una tasa de mortalidad significativamente más alta a corto plazo. La forma en que se presentan los resultados también es importante, ya que actualmente no se concibe entregar una magnitud de resultado sin tener una idea de la precisión expresada a través de la desviación estándar o el Intervalo de confianza 95%. La significación estadística sigue siendo importante, pero debe considerarse también la significación clínica de los resultados, sobre todo con los estudios con grandes números de pacientes donde es más fácil obtener la significación estadística. Los resultados del análisis de subgrupos deben ser hechos con precaución, considerando que es probable que el número de pacientes estudiados sea insuficiente y que si hay un número importante de subgrupos analizados los resultados pueden ser debidos al azar¹¹. Por otro lado, el análisis de subgrupos disminuye el tamaño muestral con la posibilidad de aumentar un error tipo II.

En relación a la aplicabilidad de los resultados, no sólo debemos considerar si la población del estudio es similar a nuestros pacientes, sino cuál es nuestra experiencia clínica, con qué recursos contamos y los valores y preferencias de los pacientes.

CONCLUSIONES

Adquirir la habilidad de analizar un ensayo clínico es fundamental para el clínico, y debiera ser preocupación fundamental de todos los programas de pregrado y de especialización médica, pues diariamente surgen nuevos estudios que proponen como útiles y beneficiosos a una amplia gama de tratamientos, pero no todos están bien realizados. Una manera de hacer este análisis es descartar rápidamente los sesgos que pueden estar presentes, siendo los más importantes los de selección y de medición. Para esto, una rápida revisión de la me-

todología de los estudios pueden ayudar a orientar respecto a la calidad de los estudios o a la validez de los resultados. Los ensayos clínicos, considerados el mejor diseño posible para estudiar causalidad, están expuestos a diversos sesgos y fallas. Cualquier tipo de sesgo puede sobreestimar significativamente los efectos del tratamiento. Los estudios pobremente diseñados hacen que los tratamientos parezcan mejores de lo que realmente son, e incluso puede parecer que funcionan, cuando en realidad no funcionan o pueden ser perjudiciales. Su pesquisa puede disminuir la exposición de los pacientes a tratamientos insuficientemente efectivos o inseguros.

REFERENCIAS

- Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. 3rd edition. Springer-Verlag, New York, 2010. p. 2.
- Marín I, Estrada MD, Casariego C. *Elaboración de Guías de Práctica Clínica en el Sistema Nacional de Salud*. Manual metodológico. Madrid; 2007.
- Schultz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-412.
- Carroll D, Tramèr M, McQuay H, et al. Randomization is important in studies with pain outcome: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. *BJA* 1996; 77: 798-803.
- Cobo E, Oliva JC, Cortés J, et al. *Ensayo clínico. Azar, riesgos de sesgo, ética*. Universidad Politécnica Catalunya, 2010.
- Ernst E, White AR. Acupuncture for back pain: A meta-analysis of randomised controlled trials. *Arch Int Med* 1998; 158: 2235-2241.
- Moher D, Ba' Pham, Jones A, Cook D. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 609-613.
- Lectura crítica de un Ensayo Clínico [internet]. [consulta el 28 de julio de 2014]. Disponible en: http://www.murciasalud.es/recursos/ficheros/136616-capitulo_10.pdf
- Moore RA, Gavaghan D, Tramèr MR, et al. Size is everything - large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. *Pain* 1998; 78: 217-220.
- Clarke M, Oxman AD, editores. *Manual del Revisor Cochrane 4.1.6* [internet]. [consulta el 28 de julio de 2014]. Acceso en: http://www.cochrane.es/files/handbookcast/Manual_Cochrane_510.pdf
- Hulley SB, Cummings SR. *Designing Clinical Research, An Epidemiologic Approach*. 2nd ed. 2001.

Correspondencia a:
Dr. David Torres P.
dtorresp@gmail.com