

Conceptos generales en bioestadística y epidemiología clínica: estudios experimentales con diseño de ensayo clínico aleatorizado

General concepts in biostatistics and clinical epidemiology: Experimental studies with randomized clinical trial design

Sebastián Estrada^a , Marcelo Arancibia^{a,b} , Jana Stojanova^{a,b} , Cristian Papuzinski^{a,b} 

^a Cátedra de Metodología de la Investigación Científica, Escuela de Medicina, Universidad de Valparaíso, Viña del Mar, Chile

^b Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Valparaíso, Chile

*Autor correspondiente cristian.papuzinski@uv.cl

Citación Estrada S, Arancibia M, Stojanova J, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Experimental studies with randomized clinical trial design. *Medwave* 2020;20(2):e7869

Doi 10.5867/medwave.2020.02.7869

Fecha de envío 7/1/2020

Fecha de aceptación 18/2/2020

Fecha de publicación 8/4/2020

Origen Este artículo es parte de una colección de “Notas metodológicas” acordada entre Medwave y la Cátedra de Metodología de la Investigación Científica de la Escuela de Medicina de la Universidad de Valparaíso

Tipo de revisión Con revisión por pares externos, por dos árbitros a doble ciego

Palabras clave clinical trial, therapeutics, relative risk, bias, epidemiology, biostatistics

Resumen

En los estudios experimentales, el investigador aplica una intervención sobre los participantes, por lo que su eje temporal es prospectivo. Estas características permiten probar relaciones causales, pero implican una rigurosa evaluación bioética y una inscripción previa del protocolo de estudio. Entre las investigaciones experimentales se encuentran los estudios preclínicos y diversos ensayos con objetivos distintos según la fase de desarrollo en que se encuentre. Entre ellos se incluyen los ensayos clínicos, cuyo objetivo principal es evaluar la eficacia y la seguridad de alguna intervención o terapia. El diseño convencional de ensayo clínico es aleatorizado, controlado y enmascarado, donde se asigna aleatoriamente a los participantes a una intervención y a sus comparadores (otra intervención, placebo o nada), lo que permite controlar el sesgo de selección y el de confusión. Asimismo, los participantes y los investigadores no conocen la intervención que fue asignada. El análisis de sus resultados debe considerar a todos los sujetos originalmente aleatorizados, lo que se conoce como análisis por intención de tratar. Otros tipos de estudios experimentales incluyen los estudios cuasiexperimentales y los estudios experimentales con controles externos. Las medidas de magnitud del efecto utilizadas son el riesgo relativo, las reducciones absolutas y relativas de riesgo y los números necesarios a tratar y dañar. Si bien los factores de confusión son controlados por el proceso de aleatorización, otros sesgos que hay que considerar son los de selección, desarrollo, detección, desgaste y reporte. En este artículo, se abordan los conceptos teóricos generales sobre los estudios experimentales en seres humanos, fundamentalmente sobre los ensayos clínicos aleatorizados, considerando sus elementos fundamentales, desarrollo histórico, aspectos bioéticos, construcción, variantes en el diseño, medidas de asociación, sesgos y pautas de reporte. Finalmente, se sintetizan algunas consideraciones que siempre hay que tener en mente durante la ejecución y la evaluación de un ensayo clínico. Esta revisión es la quinta entrega de una serie metodológica sobre conceptos generales en bioestadística y epidemiología clínica desarrollada por la Cátedra de Metodología de la Investigación Científica de la Escuela de Medicina de la Universidad de Valparaíso, Chile.

Abstract

In experimental studies, researchers apply an intervention to a group of study participants and analyze the effects over a future or “prospective” timeline. The prospective nature of these types of studies allows for the determination of causal relationships, but the interventions they are based on require rigorous bioethical evaluation, approval from an ethics committee, and registration of the study protocol prior to implementation. Experimental research includes clinical and preclinical testing of a novel intervention or therapy at different phases of development. The main objective of clinical trials is to evaluate an intervention’s efficacy and safety. Conventional clinical trials are blinded, randomized, and controlled, meaning that participants are randomly assigned to either the study intervention group or a comparator (a “control” group exposed to a placebo intervention or another non-placebo or “active” intervention—or not exposed to any intervention) to reduce selection and confounding biases,

and researchers are also unaware of the type of intervention being applied. Intention-to-treat analysis (inclusion of all originally randomized subjects) should be done. A quasi-experimental design and external controls may also be used. Metrics used to measure the magnitude of effects include relative risk, absolute and relative risk reductions, and numbers needed to treat and harm. Confounding factors are controlled by randomization. Other types of bias to consider are selection, performance, detection, and reporting. This review is the fifth of a methodological series on general concepts in biostatistics and clinical epidemiology developed by the Chair of Scientific Research Methodology at the School of Medicine, University of Valparaíso, Chile. It describes general theoretical concepts related to randomized clinical trials and other experimental studies in humans, including fundamental elements, historical development, bioethical issues, structure, design, association measures, biases, and reporting guidelines. Factors that should be considered in the execution and evaluation of a clinical trial are also covered.

Ideas clave

- Los ensayos clínicos aleatorizados evalúan la eficacia y la seguridad de una intervención controlada sobre seres humanos, permitiendo establecer causalidad.
- Existen diversos tipos de ensayos clínicos aleatorizados según su finalidad.
- Controlan eficazmente el sesgo de confusión, pero pueden presentar otro tipo de sesgos.
- Existen pautas internacionales que regulan la conducción ética y el reporte de estos estudios.

Introducción

Los estudios experimentales son aquellos en que el investigador aplica una intervención sobre los participantes, como ha sido definido en un artículo anterior de esta serie metodológica¹. Aquí se enmarcan todos los tipos de ensayo clínico que analizan intervenciones preventivas, terapéuticas, educativas, entre otras, que se realizan tanto sobre individuos como sobre grupos poblacionales^{2,3}. Algunos autores han considerado que las series de casos sin grupo control son el punto inicial de los estudios sobre intervenciones terapéuticas, ya que han aportado en el desarrollo de nuevas técnicas quirúrgicas y en el estudio de intervenciones en condiciones muy infrecuentes en donde la ejecución de un ensayo clínico se dificulta⁴. Sin embargo, los ensayos clínicos aleatorizados son el diseño metodológico de elección para evaluar la eficacia (el verdadero efecto biológico de una intervención) y la efectividad (el efecto de una intervención en la práctica clínica cotidiana)⁵, además de ser el tipo de estudio primario con la mayor capacidad de controlar los sesgos⁶.

En la década de 1920, Fisher conceptualizó la asignación aleatoria mediante la aleatorización de cultivos a recibir un tipo de semillas u otro. Más tarde, el *Medical Research Council* de Reino Unido adaptó la aleatorización a la epidemiología clínica, al realizar un ensayo clínico aleatorizado para evaluar el efecto de la estreptomina entre portadores de neumonía, publicado en 1948⁷, el que suele reconocerse como el primer ensayo clínico aleatorizado⁸. No obstante, ya en 1907 el médico William Fletcher había publicado los resultados de un ensayo clínico aleatorizado conducido para analizar el origen del beriberi en pacientes internados en un manicomio de Kuala Lumpur⁹, donde las condiciones podían ser mejor controladas. Él asignó pacientes a alimentarse con arroz blanco o integral, de acuerdo con la teoría de la época que asociaba el beriberi al consumo de arroz blanco.

Durante los últimos 70 años, los ensayos clínicos se han perfeccionado, constituyéndose en la metodología fundamental de las agencias reguladoras de medicamentos para autorizar la comercialización de estos productos¹⁰. La investigación del potencial nocivo de algunos fármacos comienza a cobrar mayor relevancia a partir de problemáticas de salud pública: muerte súbita en pacientes anestesiados con cloroformo^{11,12}, detección de aplasia de huesos largos en recién

nacidos asociada al uso de talidomida en madres con náuseas y vómitos del embarazo¹³ y, más recientemente, los efectos adversos del uso de ácido acetilsalicílico^{2,14}. Por lo tanto, diversas entidades sanitarias han decidido enfocar en la ejecución de ensayos clínicos gran parte del financiamiento disponible para la investigación en salud¹⁰. En el Ejemplo 1 se muestra un ensayo clínico aleatorizado.

Ejemplo 1. Devanand y colaboradores¹⁵ condujeron un estudio en adultos mayores con depresión y alteraciones cognitivas para evaluar la eficacia y la seguridad de la adición de donepezilo para alteraciones cognitivas. Todos los participantes (n = 81) realizaron una fase abierta del ensayo, en donde recibieron antidepresivos por 16 semanas. Luego, los pacientes se aleatorizaron a dos grupos, donde adicionalmente se les prescribió donepezilo de cinco a 10 miligramos diarios o bien placebo, durante 62 semanas más, sin ser ellos conscientes de qué se les estaba administrando. Para medir los desenlaces se les aplicó un test neuropsicológico y un cuestionario sobre actividades de la vida diaria. Los resultados no apoyaron la eficacia del donepezilo.

Este artículo corresponde a la quinta entrega de una serie metodológica de seis revisiones narrativas acerca de tópicos generales en bioestadística y epidemiología clínica, las que exploran artículos publicados disponibles en las principales bases de datos y textos de consulta especializados. La serie está orientada a la formación de estudiantes de pre y posgrado. Es realizada por la Cátedra de Metodología de la Investigación Científica de la Escuela de Medicina de la Universidad de Valparaíso, Chile. El objetivo de este manuscrito es abordar los principales conceptos teóricos y prácticos de los estudios experimentales en seres humanos, fundamentalmente los ensayos clínicos aleatorizados.

Conceptos preliminares: control, aleatorización y cegamiento

Los ensayos clínicos corresponden a diseños experimentales prospectivos (se realiza un seguimiento en el tiempo), con la capacidad de establecer relaciones de causalidad, ya que corroboran que la causa (intervención) precede al efecto (desenlace), y que esta causa corresponde a una intervención administrada por los investigadores. El que un ensayo clínico sea “controlado”, implica que se presenta

una comparación estadísticamente válida entre los resultados obtenidos en un grupo intervención y un grupo “control” o comparador. Si no se incluyera un grupo control, no podría asegurarse por qué el desenlace ocurrió⁵. Para esto, es necesario que se incluyan al menos dos grupos de pacientes y/o voluntarios sanos, cuya asignación a un tratamiento experimental o una intervención control será aleatoria; de ahí la calificación de ensayos clínicos “aleatorizados”. En el proceso de aleatorización no intervienen el investigador ni el sujeto en experimentación. La aleatorización es un fenómeno clave en este tipo de diseños, ya que es el mejor método de control de algunos de los sesgos asociados a la investigación con seres humanos. De hecho, la aleatorización ha sido considerada por algunos autores como el descubrimiento “más revolucionario y profundo de la medicina moderna”, ya que sobre ella descansan múltiples grandes descubrimientos: desde el uso de la penicilina hasta la terapia génica⁸.

Entre los ensayos clínicos controlados se encuentran aquellos abiertos o “no ciegos”, donde el participante y el investigador conocen la intervención asignada (por ejemplo, en ensayos clínicos aleatorizados que analizan intervenciones quirúrgicas). Esto fue ilustrado en el Ejemplo 1, en donde en una primera fase “abierta”, todos los participantes eran conscientes de estar usando un antidepresivo durante 16 semanas. En contraposición, los ensayos clínicos tendrán sujetos “ciegos” cuando ellos desconozcan la asignación. Tradicionalmente, se conocía como “ciego simple” a la condición en que el participante desconocía su asignación; el “doble ciego” incluía además al investigador y el “triple ciego” al evaluador. No obstante, para favorecer la claridad, hoy por hoy se solicita reportar quiénes fueron ciegos a la asignación¹⁶. Por su parte, el enmascaramiento hace referencia al mismo proceso, pero más bien al hecho de ocultar (“enmascarar”) mediante un “disfraz” el tipo de intervención recibida. Por ejemplo, si la intervención es un fármaco que se administra en un comprimido, los comprimidos de los grupos de intervención (principio activo) y comparación (placebo) tendrán características similares.

Algunos estudios han avanzado incluso más en el mantenimiento del ciego, desarrollando “placebos activos”, es decir, placebos que imitan a la intervención en algunos aspectos: por ejemplo, si un fármaco genera sequedad bucal, los participantes podrían saber si lo están recibiendo o no en función de la presencia de este efecto, por lo que el placebo podría generar también sequedad bucal pero no el efecto relacionado con el principio activo de la intervención¹⁷. En conclusión, tanto el cegamiento como el enmascaramiento guardan relación con el mismo principio^{3,18-21}. En el Ejemplo 2 se muestra un ensayo clínico aleatorizado abierto.

Ejemplo 2. Xia y colaboradores²² realizaron un ensayo abierto, en que aleatorizaron a pacientes con carcinoma hepatocelular recurrente a recibir una nueva hepatectomía o ablación percutánea por radiofrecuencia. Debido a la naturaleza de las intervenciones, no se realizó cegamiento.

Durante el reporte de los resultados suele construirse la llamada “Tabla 1”, en donde se reportan las características biosociodemográficas de interés de cada grupo, tales como sexo, edad, nivel socioeconómico, presencia de alguna patología de interés, uso de algún fármaco, entre otras. Ella tiene un valor descriptivo y analítico ya que permite comparar las diferencias basales entre los grupos.

Clasificación de los ensayos clínicos

Aunque clásicamente se ha relacionado a los ensayos clínicos con el desarrollo de fármacos, este diseño permite la evaluación de cualquier tipo de intervención. La Administración de Alimentos y Medicamentos (FDA, por sus siglas en inglés *Food and Drug Administration*), agencia reguladora estadounidense, ha presentado una clasificación de los ensayos terapéuticos que evalúan intervenciones farmacológicas, presentada a continuación. En la literatura podrá hallarse una gran mixtura en la clasificación de cada ensayo de acuerdo con esta nomenclatura, la que también se ha usado para ensayos clínicos que analizan intervenciones no farmacológicas^{2,5,10,16,23}.

- **Estudios preclínicos:** detectan problemas de seguridad del producto farmacológico, tales como carcinogenicidad y teratogenicidad. Involucran procesos como síntesis química, pruebas biológicas y estudios toxicológicos. Se realizan en animales.
- **Ensayos fase I:** estudios farmacológicos sin objetivos terapéuticos, que evalúan toxicidad, parámetros farmacocinéticos y farmacodinámicos, tolerancia, respuesta a distintas dosis y dosis máxima segura. Se realizan en voluntarios sanos y eventualmente en pacientes con enfermedad en estadios avanzados sin otra posibilidad de tratamiento. Por lo tanto, son las primeras pruebas en humanos. Usualmente son estudios abiertos y no controlados que cuentan con menos de 100 participantes.
- **Ensayos fase II:** corresponden a la primera exploración clínica del tratamiento, en donde debe definirse la posología más adecuada para los estudios fase III. Aportan información preliminar sobre eficacia y seguridad clínica. Se realizan en pacientes portadores de la enfermedad en estudio, pudiendo existir o no un grupo control, pero suelen ser ensayos clínicos aleatorizados. Por lo general incluyen entre 100 y 300 participantes. Se ha descrito una fase II temprana (II a), en donde se realizan estudios piloto para evaluar el perfil de seguridad y actividad de la nueva intervención (mayormente biodisponibilidad), y una fase II tardía (II b), que pretende garantizar aspectos como la seguridad del fármaco y la eventual superioridad de la intervención sobre otra existente. Los estudios fase II pueden actuar como un tamiz de ensayos sobre fármacos que tienen un verdadero potencial para ser evaluados en fase III.
- **Ensayos fase III:** su objetivo es demostrar el efecto de una intervención en condiciones similares a las que puedan esperarse cuando el fármaco esté disponible para su uso (estudios de confirmación terapéutica). Suelen conducirse en múltiples centros, con muestras mayores a 300 participantes (incluso miles). Se les conoce como “pivotal” o “confirmatorios”, ya que incluyen una estimación del tamaño muestral asociada a una hipótesis estadística a demostrar. Si bien los efectos tóxicos deben haber sido pesquisados en las fases I y II, en fase III se determinarán efectos secundarios frecuentes, indicando el tipo de paciente particularmente susceptible a presentarlos. Los ensayos fase III son el soporte para la autorización del registro y comercialización de un fármaco.
- **Ensayos fase IV:** evalúan el fármaco en una población distinta a la originalmente estudiada, aportando información adicional sobre riesgos, eventos adversos, beneficios, nuevos usos, efectos a largo plazo (farmacovigilancia), interacciones farmacológicas, entre otros. Se realizan después de la aprobación y comercialización del producto farmacéutico (*postmarketing*). Los estudios de fase IV corresponden a estudios observacionales (por ejemplo, series de casos, estudios de casos y controles, estudios de cohorte). En consecuencia, están sujetos a los sesgos propios de estos diseños. No obstante, proporcionan información de gran importancia sobre la aplicación del fármaco o intervención en el “mundo real”.

Indistintamente de si los ensayos clínicos estudian intervenciones farmacológicas o no farmacológicas, pueden ser catalogados como

unicéntricos en el caso de que se desarrollen por un único grupo de investigación en solo un centro hospitalario; o multicéntricos, cuando un protocolo de investigación en común es ejecutado por más de un grupo de investigación en más de un centro. Estos últimos permiten estudiar a un mayor número de participantes en menos tiempo, con conclusiones más fiables y representativas de la población. Sin embargo, su planificación, coordinación, supervisión, gestión y análisis de datos es más compleja^{16,23}.

Finalmente, es frecuente encontrar en la literatura publicada la denominación de “estudio piloto” a ciertos ensayos clínicos (Ejemplo 3)²⁴. Estos corresponden a ensayos preliminares cuyo objetivo es realizar un sondeo para ejecutar un ensayo clínico posterior de mayor relevancia. Los estudios piloto entregan luces respecto a la precisión de la hipótesis, definición de la muestra (criterios de elegibilidad) y de la intervención, estimación del tiempo requerido para el estudio e información sobre los eventuales datos perdidos. Además, de manera muy importante, aportan evidencia para la determinación del tamaño muestral del ensayo clínico posterior^{16,25}.

Ejemplo 3. Pruiksmá y colaboradores²⁴ realizaron un ensayo clínico aleatorizado piloto para obtener información preliminar sobre la eficacia, credibilidad y aceptabilidad de distintas técnicas cognitivo-conductuales para el tratamiento de militares con pesadillas relacionadas a traumas psíquicos. El estudio fue registrado en www.ClinicalTrials.gov y su identificador fue NCT02506595. Se aleatorizaron 40 participantes a dos intervenciones distintas, midiendo los desenlaces luego de cada sesión de trabajo y un mes posterior a la finalización de las terapias. Las intervenciones demostraron un efecto moderado en la reducción de las pesadillas. Los autores concluyeron que un ensayo clínico aleatorizado con un adecuado poder estadístico es necesario para confirmar los hallazgos.

Aspectos bioéticos y registro de protocolo

La Declaración de Helsinki fue desarrollada por la Asociación Médica Mundial en 1964 para ser una guía ética de la investigación con seres humanos, incluyendo aspectos como deberes de quienes investigan en seres humanos, importancia del protocolo de investigación, investigación con personas vulnerables, consideraciones sobre riesgos y beneficios, relevancia del consentimiento informado, mantenimiento de la confidencialidad e información de los hallazgos a los participantes del estudio. Aunque no es legalmente vinculante en sí misma, muchos de sus presupuestos están contenidos en las legislaciones de los países al respecto, por lo que debe considerarse en la construcción de cualquier estudio con seres humanos. La declaración cuenta con múltiples revisiones a la fecha^{17,26} (<https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/>).

Para dar inicio a un ensayo clínico, es necesario tener en mente el principio básico estipulado por la Declaración de Helsinki²⁶ en cuanto a las posibles intervenciones a estudiar: los posibles beneficios, riesgos, costos y eficacia de toda intervención nueva deben ser evaluados mediante su comparación con las mejores intervenciones probadas, excepto en las siguientes circunstancias:

1. El uso de placebo o no tratamiento es aceptable en estudios donde no exista una intervención probada.
2. Cuando por razones metodológicas científicamente sólidas y convincentes, sea necesario para determinar la eficacia y la seguridad de una intervención el uso de cualquier intervención menos eficaz que la mejor probada, el uso de un placebo o ninguna intervención.

Los ensayos clínicos requieren de un protocolo de investigación que debe ser evaluado y aprobado por un comité ético-científico (así como toda investigación con seres humanos^{1,27,28}), y registrado en algún repositorio disponible antes de que ocurra la inscripción del primer participante^{29,30}. Su finalidad es detectar cualquier desviación del protocolo original una vez realizado el estudio, garantizando que los autores reporten los desenlaces que declararon en un comienzo como clínicamente relevantes, lo que evita, por lo tanto, el informe selectivo de resultados³¹. Este proceso otorga transparencia y visibilidad a la investigación clínica, permitiendo a quienes desarrollen futuros ensayos clínicos y revisiones sistemáticas de ensayos clínicos, contar con un mapa general de la investigación en curso. Todo esto se ha conceptualizado en el modelo de Buenas Prácticas Clínicas. Se trata de un estándar para el diseño, conducción, desempeño, monitoreo, auditoría, registro, análisis y reporte de ensayos clínicos, que cautela la confiabilidad de los resultados en un marco de integridad investigativa y confidencialidad de los participantes¹⁰.

A fines del siglo XX, se originaron diversos registros públicos para ensayos clínicos. En Estados Unidos se creó el registro *Clinical Trials*, financiado con fondos públicos (<http://clinicaltrials.gov>), mientras que en Europa se estableció el registro ISRCTN (<http://isrctn.com>), reconocido por la Organización Mundial de la Salud y por el Comité Internacional de Editores de Revistas Médicas (ICMJE, por sus siglas en inglés *International Committee of Medical Journal Editors*) y respaldado por el *Medical Research Council* y el Programa de Investigación y Desarrollo del Sistema Nacional de Salud, ambas organizaciones británicas³². Por su parte, Cochrane dispone del *Cochrane Central Register of Controlled Trials* (<https://www.cochranelibrary.com/central>) y la Organización Mundial de la Salud cuenta con la Plataforma Internacional de Registro de Ensayos Clínicos (ICTRP, por sus siglas en inglés *International Clinical Trials Registry Platform*) (<https://www.who.int/ictrp/en/>).

Componentes y procedimientos esenciales en los ensayos clínicos

Reclutamiento y aleatorización de participantes

El reclutamiento de participantes para un ensayo clínico es por lo general un muestreo no probabilístico, donde se incorporan aquellos sujetos que cumplen con los criterios de elegibilidad establecidos en el protocolo del estudio. También se le conoce como “muestreo por conveniencia”. Luego, los sujetos son aleatorizados a una intervención o a su comparación, lo que, como se abordará más adelante, es el mejor método para controlar los sesgos de selección y confusión^{3,33,34}. Esta asignación no discrecional de participantes a los grupos en estudio debe realizarse estrictamente por azar para garantizar que todos los participantes tienen la misma probabilidad de ser incluidos a cualquiera de los grupos. A medida que avanza este proceso, los grupos tienden a ser más homogéneos, tanto en las variables de confusión como en otras variables que podrían afectar sobre el desenlace, pero que no podrían ser evaluadas *a priori*.

La aleatorización puede realizarse mediante una tabla de números aleatorios hallada en un libro de estadística, pero usualmente se emplean métodos informáticos de aleatorización, como una secuencia generada por un computador. Una mención particular merece el ocultamiento de la secuencia de aleatorización, la que debe ser desconocida por los investigadores y participantes del ensayo clínico, de modo de no poder predecir el grupo al que será asignado el siguiente participante incluido.

La estimación de la cantidad de participantes que deben ser aleatorizados (cálculo del tamaño de la muestra) es parte primordial de la aleatorización. ¿Cuántos participantes se requieren para igualar los factores de confusión entre los grupos de intervención? No necesariamente un número mayor será mejor, ya que podría exponerse innecesariamente a personas a los riesgos de una intervención. De este modo, si la cantidad de pacientes aleatorizados es menor que el tamaño de la muestra estimada, sus resultados estarán sesgados aun cuando haya existido la aleatorización⁸.

Entre los distintos tipos de aleatorización se encuentran la aleatorización simple, basada en una secuencia única de asignaciones aleatorias, manteniendo completamente la asignación al azar. En ensayos clínicos con grandes tamaños de muestra, la aleatorización simple puede generar números similares de participantes entre los grupos, pero en estudios que involucran pocos participantes puede resultar en cantidades desiguales en cada grupo³⁵. Otra forma de aleatorización es la aleatorización en bloques, cuyo objetivo es asegurar que los tamaños de cada grupo sean similares¹⁷. Cada bloque contiene a un número similar de participantes asignado a cada tratamiento, donde el total de participantes ha sido predeterminado por los investigadores. Luego, los bloques son asignados aleatoriamente a cada grupo.

El problema de la aleatorización en bloques es que los grupos generados pueden ser desiguales en cuanto a ciertas variables de interés³⁵. Ante esto, la aleatorización estratificada se aplica para garantizar que se asignará a cada grupo una cantidad similar de participantes con una característica de importancia para el estudio, las que deben identificarse por los investigadores. En este tipo de aleatorización, se configuran distintos bloques de participantes con combinaciones de covariables que puedan influir en la variable dependiente que se quiere explicar (aleatorización según factores pronósticos). Luego, se realiza una aleatorización simple dentro de cada estrato para asignar a los sujetos a alguno de los grupos de intervención. Por lo tanto, se puede intuir que para llevar a cabo una aleatorización estratificada, es necesario conocer con exactitud las características de cada sujeto^{17,35,36} (Ejemplo 4).

Finalmente, un método de aleatorización que se ha empleado en los ensayos clínicos con un tamaño de muestra pequeño es la aleatorización adaptativa, ya que en ella se asigna secuencialmente un nuevo participante a un grupo de intervención en particular, considerando las asignaciones previas de los participantes además de las covariables específicas. La aleatorización adaptativa usa el método de minimización, evaluando el desbalance del tamaño muestral entre múltiples covariables, lo que podría ocurrir al aplicar una aleatorización simple en un ensayo clínico con una muestra reducida^{35,37,38}.

Ejemplo 4. Burns y colaboradores³⁹ diseñaron un ensayo clínico aleatorizado con grupos paralelos de fase III, para evaluar la eficacia y la seguridad de pioglitazona en dosis bajas para retrasar el inicio de deterioro cognitivo leve asociado a la enfermedad de Alzheimer. Mediante el algoritmo de clasificación “BRAA”, que considera ciertos genotipos y la edad de la persona, se estratificó a los participantes en un grupo de alto riesgo y en otro de bajo riesgo de desarrollar deterioro cognitivo leve debido a enfermedad de Alzheimer. El estrato de alto riesgo se aleatorizó a recibir pioglitazona en dosis baja o placebo.

Cualquier asignación que no responda al azar, tal como la asignación por alternancia (por ejemplo, el primer participante es asignado al

grupo intervención, el segundo al grupo control y así sucesivamente), de acuerdo al día de la semana, según la letra inicial del primer apellido, entre otros, no puede calificarse como una asignación aleatoria⁴.

Medición de resultados

Idealmente, deben aplicarse métodos de enmascaramiento que aseguren el ciego de los participantes, como fue previamente señalado. El ciego permite que el énfasis que pongan los encargados del ensayo en la medición de resultados será el mismo para todos los grupos y que los pacientes no tendrán la influencia de saber si están en el grupo intervención o no, lo que disminuiría las respuestas subjetivas al tratamiento. Por tanto, el ciego previene la ocurrencia del sesgo de medición.

Tres fenómenos relacionados que pueden suceder a nivel de los participantes son los siguientes: el efecto placebo^{40,41}, el efecto nocebo⁴⁰ y el efecto Hawthorne^{42,43}. El primero se asocia a una mejoría reportada por el participante después de recibir una sustancia sin desenlaces relevantes demostrados, esto es, un placebo. Por el contrario, en el efecto nocebo, una sustancia o intervención sin efectos médicos empeora el estado de salud de una persona debido a las creencias negativas que el participante tiene sobre ella. El efecto Hawthorne, también conocido como “efecto del observador”, ocurre cuando los participantes de un ensayo clínico modifican su conducta habitual al saber que están siendo observados por un tercero, lo que impacta en los efectos de la intervención. Estos tres fenómenos son más prominentes en ensayos clínicos aleatorizados que analizan desenlaces reportados por los participantes, es decir, donde existe un componente subjetivo mayor (Ejemplo 5).

Ejemplo 5. Hodgins y colaboradores⁴⁴ realizaron un ensayo clínico en el que compararon el efecto del antagonista del receptor tipo 1 del factor liberador de corticotropina (CRHR1) versus placebo en mujeres con trastorno de estrés postraumático. Para evaluar el estado basal y postintervención se utilizaron escalas de autoreporte y entrevistas clínicas estructuradas orientadas a medir síntomas de estrés postraumático y depresión, además de pruebas neurocognitivas y de funcionamiento general. El grupo placebo demostró tener un efecto sustancial y no inferior al fármaco en cuanto a los desenlaces reportados por los participantes.

Análisis de datos

Una vez medidos los desenlaces de interés tras el tiempo necesario, es esencial el poder incluir en el análisis estadístico primario a todos (o la mayor parte de) los participantes aleatorizados, ya que por algún motivo, algunos sujetos podrían abandonar el tratamiento, lo que debe ser evitado por los investigadores. El análisis primario debe realizarse incluyendo y analizando a todos los pacientes como parte del grupo original al que fueron asignados, lo que se conoce como análisis por intención de tratar. El principio del análisis por intención de tratar es mantener hasta el final del estudio el objetivo logrado con la aleatorización, es decir, el balance de los factores pronósticos conocidos y desconocidos y de los factores de confusión, disminuyendo la probabilidad de sesgar los resultados⁴⁵.

No obstante, este enfoque se ve dificultado cuando existe una baja adherencia al tratamiento asignado, o cuando el ensayo no logra el seguimiento de todos los pacientes. Algunos investigadores resuel-

ven este problema excluyendo del análisis a los pacientes que no adherieron al protocolo, lo que se conoce como análisis por protocolo. Este solo incorpora a participantes que recibieron la intervención asignada por la aleatorización, lo que en un comienzo puede parecer razonable, pues interesa conocer los efectos de la intervención en los pacientes que realmente la recibieron. Sin embargo, es reconocido que la adherencia habitual a cualquier tratamiento suele ser inferior a la estimada teóricamente, por lo que el análisis por protocolo no representaría lo que sucede en la realidad⁴⁶. A su vez, se ha demostrado que los pacientes menos adherentes a una terapia, aunque esta sea un placebo, tienden a tener peor pronóstico que los que sí adhieren a ella⁴⁷. Al excluirlos del análisis, se está estudiando los pacientes de mejor pronóstico, efecto que se magnifica si se decidiera incluir al grupo de participantes que no recibió la intervención (y que podrían considerarse como semejantes al grupo control) en el grupo control, perdiendo así el efecto de la aleatorización⁴⁸ (Ejemplo 6).

Ejemplo 6. Un estudio evaluó la utilidad de una técnica de bloqueo lumbar comparada con analgesia para disminuir el dolor lumbar, según criterios de dolor de la Organización Mundial de la Salud. Se aleatorizaron 120 pacientes, 60 a cada grupo, pero dado que 20 pacientes aleatorizados a la técnica de bloqueo no califican desde el punto de vista anestésico para someterse a una cirugía, reciben un analgésico según protocolo. Al analizar los datos en virtud de la intervención que recibieron (análisis por protocolo), se puede incurrir en sobrestimar el efecto del bloqueo lumbar, dado que los pacientes más graves se analizaron en el grupo con analgésicos, pudiendo reportar este grupo también más eventos adversos.

Tipos de estudios experimentales en seres humanos

En este artículo se han abordado mayormente los conceptos relacionados con el ensayo clínico aleatorizado clásico, cuyas variantes y especificaciones se describen a continuación. No obstante, existen otros diseños que también estudian los efectos de una intervención sobre los participantes.

Ensayo clínico aleatorizado clásico

Corresponde al ensayo fase III de la clasificación anterior. Se aleatorizan los participantes a recibir una intervención de prueba o una comparación. La aleatorización garantizaría que las características de los participantes se distribuyan homogéneamente entre los grupos, por lo tanto, cualquier diferencia significativa en el desenlace entre los grupos puede ser atribuida a la intervención y no a otro factor no identificado. Si el ensayo clínico busca evaluar una intervención en condiciones ideales (no cotidianas) y controladas de manera rigurosa, se trata de un ensayo clínico explicativo (estudios de eficacia). Mientras que, si su evaluación se da en un contexto que emula las circunstancias de la vida o la práctica clínica cotidiana, recibe el nombre de ensayo clínico pragmático⁴⁹. En este sentido, se han desarrollado herramientas para evaluar el nivel de pragmatismo de un ensayo clínico⁵⁰.

Otra denominación se relaciona con el objetivo del estudio, ya que existen ensayos clínicos aleatorizados cuyo objetivo es demostrar la superioridad, equivalencia o no inferioridad de una intervención sobre otra⁵¹. En el primer caso, se busca probar que un tratamiento es mejor que otro. En el segundo, será apropiado conducir un estudio de equivalencia si el nuevo tratamiento tuviera similar eficacia, pero

ciertos beneficios como menos efectos adversos, un uso más simple o un menor costo económico. Los estudios de no inferioridad pueden considerarse como un caso especial de los estudios de equivalencia. Estos solo analizan que una intervención no es peor que otra ya existente¹⁷.

En virtud del diseño del estudio encontraremos el ensayo clínico aleatorizado paralelo. Es el modelo más frecuente, en donde cada grupo de participantes recibe una intervención simultáneamente. Por su parte, en los ensayos clínicos aleatorizados cruzados (*crossover*), cada participante recibe consecutivamente cada intervención en estudio, de modo de que cada sujeto es su propio comparador. Por lo tanto, requieren de un tamaño de muestra menor, ya que los sujetos aleatorizados inicialmente al grupo de intervención luego recibirán la comparación y viceversa, por lo que sumarán participantes a ambos grupos. Una desventaja de este diseño es el fenómeno de *carryover*, en donde los efectos de la primera intervención pueden interferir con los efectos de la segunda. Por ende, este tipo de ensayos es útil en intervenciones que duran un periodo corto. De todos modos, es conveniente espaciar temporalmente ambas intervenciones (periodo de lavado), en orden a disminuir la probabilidad de que la primera interfiera con la segunda^{52,53}. Por lo tanto, al suspender la intervención, la condición del sujeto que la recibió debe ser la misma que antes de recibirla; ya que si cambia, la segunda intervención se aplicaría sobre un participante distinto al que recibió la primera. Es por esta razón que este diseño se limita al estudio de condiciones crónicas y estables en el tiempo⁵⁴ (Ejemplo 7).

Ejemplo 7. Mason y colaboradores⁵⁵ desarrollaron un ensayo clínico aleatorizado cruzado en portadores de diabetes mellitus tipo 2 (26 hombres y cinco mujeres). Los participantes fueron aleatorizados a recibir suplementación con ácido ascórbico (un gramo diario) o placebo, evaluando los efectos sobre la glicemia postprandial y la presión arterial. Cada grupo de participantes recibió alguna de las intervenciones durante cuatro meses y, luego de un periodo de lavado, recibieron por otros cuatro meses la otra intervención. Los autores concluyeron que el uso de ácido ascórbico es un complemento efectivo de la terapia anti diabética para mejorar el control glicémico y de la presión arterial.

Otro diseño existente es el ensayo clínico factorial, el que permite responder conjuntamente a dos o más preguntas de investigación. Un ejemplo es cuando dos o más intervenciones son evaluadas separadamente y en combinación contra un control. Los participantes son aleatorizados dos o más veces a alguno de los grupos de intervención, dependiendo del número de terapias a estudiar. La gran ventaja es que aporta más información respecto a un estudio con diseño en paralelo. Además, permitiría evaluar la interacción entre dos tratamientos⁵⁶.

Otra modalidad de ensayo clínico aleatorizado es aquella en que se aleatoriza por conglomerados (*clusters*). Vale decir, se asignan aleatoriamente grupos de participantes (por ejemplo, centros de salud, áreas geográficas) a una intervención o su comparación. Son útiles en el estudio de individuos con características biológicas o psicosociales similares y cuando la intervención analizada tiene un efecto grupal, tales como intervenciones no farmacológicas en políticas públicas comunitarias^{57,58} o el efecto de una vacuna⁵.

Finalmente, existen ensayos clínicos aleatorizados de discontinuación. En ellos, los pacientes que ya reciben algún tratamiento son aleatorizados a continuar con su terapia o discontinuarla y recibir un

placebo. Los diseños de discontinuación se aplican en terapias crónicas no curativas cuyo efecto es de pequeña magnitud⁵⁴.

Estudios cuasi-experimentales

Se caracterizan por no presentar un proceso de aleatorización de los participantes al grupo intervención o al grupo comparador, por lo que también se conocen como ensayos controlados no aleatorizados. Si bien existen numerosos ejemplos de tipos de estudios cuasiexperimentales en la investigación biomédica, este diseño proviene del campo de la psicología y las ciencias sociales⁵⁹, donde se ha considerado que en ciertas circunstancias no es posible la asignación aleatoria de sujetos a condiciones experimentales. Existen diversos tipos de estudios cuasiexperimentales; entre ellos, los diseños antes/después o pretest/postest y las series de tiempo interrumpidas. Destacaremos el diseño de antes-después, en el cual se mide una misma variable de manera previa y posterior a una intervención (cada participante actúa como su propio control). Sus resultados son útiles cuando el efecto es de gran magnitud, se observa consistentemente en la mayoría de los participantes y, por ende, es improbable que se explique por efecto del azar⁴ (Ejemplo 8).

Ejemplo 8. Se ejecutó un estudio antes/después para evaluar el efecto de la amigdalectomía sobre la calidad de vida en niños con apnea del sueño. Para ello, se administró un instrumento para medir la calidad de vida en niños con indicación de amigdalectomía un mes antes y un mes después de la intervención. Se demostró que la calidad de vida aumentó significativamente luego de la intervención.

Los estudios cuasiexperimentales suelen ser más simples e implican un menor costo económico que los ensayos clínicos aleatorizados. Es por esto que constituyen una opción cuando la asignación aleatoria es poco factible de realizar, cuando existe un impedimento bioético o cuando se requiere realizar la intervención en condiciones naturales. Sus desventajas se asocian con una gran susceptibilidad a los sesgos de confusión y selección. Además, el efecto placebo y el efecto Hawthorne son especialmente relevantes, lo que podría aminorarse si es que los sujetos no son conscientes de la intervención de la que participan^{60,61}. Durante el análisis de los datos deben considerarse idealmente métodos estadísticos para grupos pareados, ya que se realizarán mediciones repetidas en el tiempo a un mismo grupo de sujetos⁶². Por todo lo anterior, la interpretación y extrapolación de los datos obtenidos desde los estudios no aleatorizados es compleja¹⁷.

Estudios experimentales con controles externos o históricos

Utilizan como grupo control a sujetos que no provienen de la misma población desde la que se obtuvo la muestra (muestra no concurrente o no contemporánea). Por lo tanto, la comparación se realizará a partir de datos de pacientes ya publicados o desde registros de una institución de salud, es decir, con personas que ya han recibido un tratamiento y una evaluación. Una desventaja de esta metodología son las diferencias que puedan existir entre el grupo intervención y el grupo control, haciéndolos poco comparables⁴, ya que existen diferencias en el contexto temporal en que se aplicaron los tratamientos, las personas que realizaron la intervención, entre otros¹⁷.

Medidas de asociación

Debido a que se trata de un diseño prospectivo, la medida de asociación a usar será el riesgo relativo (RR), el que se entiende como la

razón de los riesgos absolutos entre el grupo de individuos expuestos a la intervención y los no expuestos. Si el riesgo relativo es igual a 1, se asume que no existe asociación entre la intervención y el desenlace (esto es, el intervalo de confianza asociado no incorpora el valor 1). Si es mayor a 1, la intervención aumenta la probabilidad de que suceda el desenlace; mientras que, si es menor a 1, la intervención disminuye la probabilidad de ocurrencia del desenlace. La interpretación será distinta en virtud del desenlace medido, ya que este puede ser favorable o desfavorable. Por lo tanto, si el resultado estudiado es disminución de ánimo depresivo, un riesgo relativo mayor a uno será favorable, mientras que si se mide mortalidad, un riesgo relativo mayor a uno será desfavorable⁶³⁻⁶⁵.

$$RR = \frac{R(\text{expuestos})}{R(\text{no expuestos})}$$

Otro modo para expresar la magnitud de asociación es mediante diferencias. Una de ellas es la reducción absoluta de riesgo (RAR), riesgo atribuible o reducción de riesgo, que corresponde a la diferencia entre los riesgos del grupo control (no expuesto) e intervención (expuesto). Esto quiere decir que la reducción de riesgo es atribuida a la intervención. Si su resultado es negativo, se interpreta como un aumento absoluto del riesgo⁶³⁻⁶⁵.

$$RAR = R(\text{no expuestos}) - R(\text{expuestos})$$

Por su parte, la reducción relativa de riesgo (RRR) corresponde a la diferencia de riesgo entre los dos grupos respecto al grupo control, o cociente entre la reducción absoluta de riesgo y el riesgo del grupo control. Si su resultado es negativo, se interpreta como un aumento relativo del riesgo. La reducción relativa del riesgo frecuentemente sobreestima los efectos del tratamiento, por lo que la reducción absoluta de riesgo debería reportarse cuando sea posible⁶³⁻⁶⁵.

$$RRR = \frac{RAR}{R(\text{no expuestos})} = 1 - RR$$

Una medida derivada de la reducción absoluta de riesgo es el número necesario a tratar (NNT), la que cuantifica el número de pacientes a tratar para evitar que ocurra un evento⁶³⁻⁶⁵. Por ejemplo, si el número necesario a tratar para un antibiótico fuese de 15, se interpreta que se debe tratar a 15 pacientes para la recuperación de un cuadro infeccioso.

$$NNT = \frac{1}{RAR} \cdot 100$$

Por otra parte, el número necesario a dañar corresponde a un índice de los eventos adversos asociados a un tratamiento, significando el número de pacientes que debería recibir un tratamiento en lugar de otro para que un paciente adicional presente un evento perjudicial. Cuando el resultado del número necesario a tratar es negativo, debe interpretarse como el número necesario a dañar⁶³⁻⁶⁵.

Si bien la reducción absoluta del riesgo, la reducción relativa del riesgo y el número necesario a tratar otorgan información sobre la magnitud del efecto del tratamiento, están fuertemente relacionadas a la variabilidad del parámetro medido y al tamaño de la muestra. Por ello, deben reportarse junto a intervalos de confianza que entreguen información sobre la precisión de los hallazgos⁶⁶.

Sesgos

Un sesgo presente en cualquier tipo de investigación es el sesgo de confusión, ya que nunca podrá eliminarse del todo. Sin embargo, los ensayos clínicos logran reducirlo significativamente gracias al proceso de aleatorización, el que permite la distribución homogénea de las variables conocidas y desconocidas entre los grupos de estudio. A continuación, se nombran algunos de los sesgos observados en los ensayos clínicos.

Sesgo de selección

Ocurre cuando los métodos para seleccionar la muestra desde la población favorecen a un grupo respecto al otro. También aparece cuando no se incluye a alguna proporción relevante de la población objetivo. Este sesgo es controlado con un adecuado proceso de selección de participantes y con su posterior aleatorización⁶⁷.

Sesgo de desarrollo (*performance bias*)

Existe una diferencia sistemática entre grupos respecto al cuidado y seguimiento entregado. Por ejemplo, está presente cuando los investigadores llevan un seguimiento más estricto de los pacientes asignados a la intervención bajo estudio. La estandarización de los procedimientos, el entrenamiento adecuado del personal y el enmascaramiento combaten esta fuente de error sistemático^{67,68}.

Sesgo de detección o información

Sucede cuando el desenlace en estudio es “detectado” diferencialmente entre los grupos, lo que puede conducir a resultados diferentes. El sesgo de detección es más prominente durante el registro de desenlaces subjetivos reportados por los participantes (por ejemplo, respuesta analgésica). De esta forma, si un investigador registra los resultados observados de un modo que avale su creencia, emergerá el sesgo de detección. Este sesgo se controla mediante el enmascaramiento^{67,68}.

Sesgo de desgaste o post aleatorización (*attrition bias*)

Se presenta cuando existen diferencias sistemáticas en el seguimiento de los participantes del ensayo clínico, ocasionando por ejemplo pérdidas de seguimiento, las que incrementan la incertidumbre en los resultados. Este sesgo se controla realizando un análisis por intención de tratar^{67,68}.

Sesgo de reporte (*reporting bias*)

Este sesgo se pesquisa durante la presentación de los resultados, donde puede ocurrir un reporte selectivo de los desenlaces de mayor interés o de aquellos que demuestran la hipótesis en estudio, lo que aumentaría el impacto del estudio y la probabilidad de que sea publicado. El sesgo de reporte puede evaluarse mediante la revisión del registro inicial del ensayo clínico y/o de su protocolo publicado para luego compararlo con lo que se ha reportado^{67,68}.

Pautas de reporte

El adecuado reporte de los ensayos clínicos es de gran importancia, pues permite comprender e interpretar los resultados y las conclusiones, así como garantizar su reproducibilidad⁶⁹. Con el fin de estandarizar criterios en el reporte de los ensayos clínicos y facilitar la lectura crítica y la interpretación, aparece a mediados de los noventa la iniciativa *Consolidated Standards of Reporting Trials* (CONSORT, por sus siglas en inglés)⁷⁰, la que está en constante revisión, actualización

y especialización. Esta propuesta está constituida por 25 ítems agrupados en seis dominios: título y resumen, introducción, metodología, resultados, discusión y otra información, debiendo agregarse además un diagrama de flujo de los participantes en el ensayo, lo que permite una rápida comprensión de los aspectos más relevantes de la ejecución del estudio. Existen múltiples versiones de *Consolidated Standards of Reporting Trials* adaptadas para intervenciones en específico: para estudios de no inferioridad⁷¹, para estudios piloto⁷², para estudios con intervenciones herbales⁷³, para ensayos clínicos pragmáticos⁷⁴, para ensayos clínicos con intervenciones psicológicas y sociales⁷⁵, para ensayos clínicos de participante único⁷⁶, para ensayos clínicos cruzados⁵³, entre otras. Brevemente, en la lectura de un ensayo clínico publicado deben responderse las siguientes preguntas⁵:

- ¿Se trata de un ensayo clínico de alta calidad que aborda una pregunta importante?
- ¿La aleatorización se realizó adecuadamente?
- ¿Cuán completo fue el seguimiento? ¿Fue similar entre los grupos del estudio?
- ¿Se buscaron desenlaces positivos y negativos? ¿Fueron ciegas esas evaluaciones?
- ¿Los resultados se aplican a la práctica clínica?
- ¿Los pacientes fueron adecuadamente descritos?
- ¿La intervención fue apropiadamente descrita?
- ¿Se realizó un análisis por intención de tratar o por protocolo?

Por otra parte, la iniciativa *Transparent Reporting of Evaluations with Nonrandomized Designs* (TREND, por sus siglas en inglés)⁷⁷, está destinada al reporte de estudios de intervención no aleatorizados. Consta de 22 ítems agrupados en cinco dominios: título y resumen, introducción, métodos, resultados y discusión.

Paralelamente, para evaluar la calidad de los ensayos clínicos se han presentado escalas como la de Jadad⁷⁸ y la desarrollada por Cochran⁷⁹.

Consideraciones finales

Usualmente, la primera tabla de resultados que se reporta en las publicaciones de ensayos clínicos es la “Tabla 1”, donde se describen condiciones de interés de los participantes. Es igualmente frecuente encontrar en la literatura publicada que esta “Tabla 1” presente los valores p resultantes de la comparación con una prueba estadística de las características entre el grupo de intervención y el grupo comparador, siendo esperable que estos valores p sean mayores al nivel de significancia estadística. Ello demostraría que los grupos no difieren en sus características biosociodemográficas y, por lo tanto, ellas no se comportarían como variables de confusión.

Sin embargo, esto no es metodológicamente correcto, ya que la hipótesis estadística subyacente al ensayo clínico y asociada a la pregunta de investigación se relaciona con la estimación de una diferencia entre dos intervenciones y no con la evaluación de la frecuencia de una variable entre grupos. Este hecho ha sido enfatizado por la guía *Consolidated Standards of Reporting Trials*^{70,80}. Algunos autores han dispuesto una solución: si existe una diferencia mayor al 10% entre la frecuencia de las variables estudiadas, podrá suponerse que existe un potencial efecto de confusión. Si esta diferencia existe, quiere decir que la aleatorización no controló todo el efecto de la confusión, por lo que los resultados deben evaluarse teniendo en mente este efecto residual o, mejor aun, realizar un modelo de regresión multivariado que incorpore a la variable en cuestión para evaluar su efecto como factor de confusión⁸.

El carácter prospectivo de los ensayos clínicos permite encontrar relaciones causales y adoptar medidas para asegurar la calidad de los datos obtenidos. No obstante, es necesario que los investigadores esperen el tiempo suficiente (a veces bastante tiempo) para que los desenlaces de interés ocurran. En este punto vale la pena reconocer el análisis de supervivencia, método estadístico mayormente usado en los estudios de cohorte²⁸, que permite analizar el tiempo en que ocurre un desenlace esperado y comparar las curvas temporales entre distintos grupos. Otro método estadístico aplicado en los ensayos clínicos es el análisis secuencial, el que consiste en la realización de análisis intermedios para evaluar la necesidad de continuar o detener un ensayo en virtud de la demostración de la hipótesis y el balance costo-beneficio, obedeciendo a ciertas reglas de detención. El análisis secuencial debe especificarse en el protocolo del estudio^{81,82}.

Algunas críticas que se han hecho a los ensayos clínicos aleatorizados se relacionan con la poca representatividad que puede tener la muestra de participantes respecto al real espectro en la práctica clínica, cuyo muestreo es no probabilístico y se realiza sobre personas seleccionadas con criterios de elegibilidad rigurosos. Asimismo, la estandarización de la intervención tampoco se asemeja a lo que sucede realmente, en donde las intervenciones serían menos controladas y más heterogéneas. Por esto, la validez externa (extrapolación de resultados) debe evaluarse con cautela, ya que los resultados del ensayo clínico pueden distar de lo observado en la práctica clínica, en donde intervienen múltiples factores que no fueron controlados en el estudio².

No todas las interrogantes pueden ser respondidas mediante un estudio de experimental, como por ejemplo, el análisis de los factores de riesgo para el desarrollo de neoplasia pulmonar, donde no sería ético que un grupo de personas recibieran una intervención deletérea para su salud. En estos casos cobran importancia los estudios observacionales²⁸. El estudio de los eventos adversos asociados al uso de fármacos debe reservarse para estudios observacionales o estudios experimentales de fases tempranas, pero no en experimentos clínicos con seres humanos. De esta forma, el análisis de eventos adversos de baja ocurrencia permanece como un desafío, puesto que los estudios disponibles cuentan en general con una baja potencia (vinculada al bajo tamaño de muestra), se requieren grupos de comparación muy similares y el análisis de la relación causa-efecto es complejo².

Es necesario diferenciar los desenlaces clínicamente relevantes o significativos de los desenlaces subrogados, intermedios, indirectos o sustitutos⁸³. Los desenlaces clínicamente relevantes orientan respecto al uso de una intervención y dan luces sobre el efecto de la terapia, es decir, la medida directa de cómo un paciente se siente, vive y funciona. Por su parte, los desenlaces subrogados pueden corresponder a un parámetro de laboratorio o un signo físico que no miden directamente el beneficio clínico central de la intervención. Así, los primeros se asocian a la persona en sí misma, mientras que los segundos a la fisiopatología de la enfermedad. Esta consideración es importante al leer ensayos clínicos aleatorizados, pues podría ocurrir que un estudio reporte múltiples desenlaces subrogados positivos pero no resultados clínicamente relevantes.

Finalmente, aunque los ensayos clínicos aleatorizados son la piedra angular para estudiar la eficacia y seguridad de una terapia, una revisión sistemática que metaanalice los resultados de múltiples ensayos clínicos que estudiaron una misma intervención, reportará un nivel de evidencia aun mayor, ya que ofrece un estimador combinado del efecto de todos los estudios primarios que incluyó⁸⁴. En cualquier

sentido, los resultados obtenidos por un ensayo clínico deben publicarse, de lo contrario, se contribuiría al sesgo de publicación al publicarse solo los resultados que son positivos a la hipótesis establecida en el protocolo. Esto favorece que la comunidad científica observe un efecto magnificado de una intervención⁶, lo que impacta negativamente en la sociedad. Se trata de un hecho contraproducente con el bien interno de la investigación, esto es, el bien social.

Notas

Roles y contribuciones de los autores

Todos los autores contribuyeron en la planificación y escritura del manuscrito original, así como en la redacción de la introducción, conceptos preliminares, ejemplos y conclusiones del artículo. SE, MA y CP: aportaron el desarrollo de los acápites clasificación, aspectos bioéticos y registro y componentes y procedimientos. SE, MA, JS y CP: elaboraron las secciones medición de resultados, análisis de datos, tipos de estudios experimentales y medidas de asociación. MA, JS y CP: realizaron los acápites sesgos y pautas de reporte. SE, MA y CP realizaron la Figura 1.

Financiamiento

Los autores declaran que no hubo fuentes externas de financiamiento.

Conflictos de intereses

Los autores completaron la declaración de conflictos de interés de ICMJE y declararon que no recibieron fondos por la realización de este artículo; no tienen relaciones financieras con organizaciones que puedan tener interés en el artículo publicado en los últimos tres años y no tienen otras relaciones o actividades que puedan influenciar en la publicación del artículo.

Aspectos éticos

Este estudio no requirió la evaluación por parte de un comité de ética-científica, debido a que trabajó sobre fuentes secundarias.

Referencias

1. Cataldo R, Arancibia M, Stojanova J, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Observational studies with cross-sectional and ecological designs. *Medwave*. 2019 Sep 25;19(8):e7698. | CrossRef | PubMed |
2. Araujo M. Categorías generales de estudios clínicos. *Medwave*. 2011;11(2):1-4. | CrossRef |
3. Friedman LM, Furberg CD, DeMets DL. Introduction to clinical trials. En: *Fundamentals of Clinical Trials*. 3a Ed. New York: Springer-Verlag; 1998: 1-12.
4. Araujo M. Estudios sobre el tratamiento de las enfermedades. *Medwave*. 2011;11(8):1-5. | CrossRef |
5. Ward H, Toledano MB, Shaddick G, Davies B, Elliott P. Types of study. En: *Oxford Handbook of Epidemiology for clinicians*. 1a Ed. Oxford: Oxford University Press; 2012: 145-196.
6. Jiménez Paneque R, Bachelet VC. Clinical trials and study reporting - CONSORT guidelines at the forefront. *Medwave*. 2015 Sep 30;15(8):e6273. | CrossRef | PubMed |
7. Group BMJP. Streptomycin treatment of pulmonary tuberculosis: a medical research council investigation. *Br Med J*. 1948 Oct;2(4582):769-82. | CrossRef |
8. Ghaemi SN. Randomization. En: *A clinician's guide to statistics and epidemiology in mental health: measuring truth and uncertainty*. 1a Ed. Cambridge: Cambridge University Press; 2009: 21-26.
9. Fletcher W. Rice and beri-beri: Preliminary report on an experiment conducted at the Kuala Lumpur lunatic asylum. *The Lancet*. 1907 Jun;169(4374):1776-9. | CrossRef |
10. Valenzuela M, Urquidí C, Martínez C. Manual para la estandarización de estudios clínicos en Chile: desde la concepción hasta la publicación. Santiago: Facultad de Medicina Universidad de Los Andes; 2017. [On line]. | Link |

11. Roberts WA. Sudden Death in a Case in Which Chloroform Was Being Administered. *Edinb Med J*. 1855 Dec;1(6):524-526. | PubMed |
12. Hutcheon DE. Sudden deaths during chloroform and cyclopropane anaesthesia. *Fed Proc*. 1948 Mar;7(1 Pt 1):229. | PubMed |
13. Somers GS. Thalidomide and congenital abnormalities. *Lancet*. 1962 Apr 28;1(7235):912-3. | PubMed |
14. Stevenson DD. Diagnosis, prevention, and treatment of adverse reactions to aspirin and nonsteroidal anti-inflammatory drugs. *J Allergy Clin Immunol*. 1984 Oct;74(4 Pt 2):617-22. | PubMed |
15. Devanand DP, Pelton GH, D'Antonio K, Ciarleglio A, Scodes J, Andrews H, et al. Donepezil Treatment in Patients With Depression and Cognitive Impairment on Stable Antidepressant Treatment: A Randomized Controlled Trial. *Am J Geriatr Psychiatry*. 2018 Oct;26(10):1050-1060. | CrossRef | PubMed |
16. Idoate A, Idoipe Á. Investigación y ensayos clínicos. En: *Farmacia Hospitalaria*. Madrid; 2002. [On line]. | Link |
17. Peacock J, Peacock P. Research design. En: *Oxford Handbook of Medical Statistics*. 1a Ed. Oxford: Oxford University Press; 2011: 1-73.
18. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet*. 2002 Feb 23;359(9307):696-700. | PubMed |
19. Morris D, Fraser S, Wormald R. Masking is better than blinding. *BMJ*. 2007;334(7597):799. | CrossRef |
20. Gálvez-Olortegui J, Gonzales-Saldaña J, García-Gómez I, Silva-Ocas I, Gutiérrez-Arana J, Gálvez-Olortegui T, et al. Control de sesgos en los ensayos clínicos: enmascaramiento o cegamiento. *Medwave*. 2015;15(11):e6349. | CrossRef |
21. Schulz KF, Altman DG, Moher D. Blinding is better than masking. *BMJ*. 2007;334(7600):918. | CrossRef |
22. Xia Y, Li J, Liu G, Wang K, Qian G, Lu Z, et al. Long-term Effects of Repeat Hepatectomy vs Percutaneous Radiofrequency Ablation Among Patients With Recurrent Hepatocellular Carcinoma: A Randomized Clinical Trial. *JAMA Oncol*. 2019 Nov 27. | CrossRef | PubMed |
23. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. *Acta Obstet Gynecol Scand*. 2018 Apr;97(4):380-387. | CrossRef | PubMed |
24. Pruikma KE, Taylor DJ, Mintz J, Nicholson KL, Rodgers M, Young-McCaughan S, et al. A pilot randomized controlled trial of cognitive behavioral treatment for trauma-related nightmares in active duty military personnel. *J Clin Sleep Med*. 2020 Jan 15;16(1):29-40. | CrossRef | PubMed |
25. Julious SA. Pilot Studies in clinical research. *Stat Methods Med Res*. 2016 Jun;25(3):995-6. | CrossRef | PubMed |
26. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013 Nov 27;310(20):2191-4. | CrossRef | PubMed |
27. Martínez D, Papuzinski C, Stojanova J, Arancibia M. General concepts in biostatistics and clinical epidemiology: observational studies with case-control design. *Medwave*. 2019 Nov 7;19(10):e7716. | CrossRef | PubMed |
28. Lazcano G, Papuzinski C, Madrid E, Arancibia M. General concepts in biostatistics and clinical epidemiology: observational studies with cohort design. *Medwave*. 2019 Dec 16;19(11):e7748. | CrossRef | PubMed |
29. De Angelis C, Drazen J, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med*. 2004;351(12):1250-1. | CrossRef |
30. Reveiz L, Villanueva E, Iko C, Simera I. Compliance with clinical trial registration and reporting guidelines by Latin American and Caribbean journals. *Cad Saude Publica*. 2013 Jun;29(6):1095-100. | PubMed |
31. Isojarvi J, Wood H, Lefebvre C, Glanville J. Challenges of identifying unpublished data from clinical trials: Getting the best out of clinical trials registers and other novel sources. *Res Synth Methods*. 2018 Dec;9(4):561-578. | CrossRef | PubMed |
32. Tonks A. A clinical trials register for Europe. *BMJ*. 2002 Dec 7;325(7376):1314-5. | PubMed |
33. Araujo Alonso M. Confusion in clinical studies. *Medwave*. 2012;12(4):e5349. | CrossRef |
34. Ruiz J. Variables y fuentes de error en la medición (confusión, sesgo y azar). En: *Epidemiología clínica: investigación clínica aplicada*. 2da Ed. Bogotá: Editorial Médica Panamericana; 2015.
35. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011 Jan;4(1):8-11. | CrossRef | PubMed |
36. Kundt G. Comparative evaluation of balancing properties of stratified randomization procedures. *Methods Inf Med*. 2009;48(2):129-34. | CrossRef | PubMed |
37. Fonarow GC. Randomization-There Is No Substitute. *JAMA Cardiol*. 2016 Sep 1;1(6):633-5. | CrossRef | PubMed |
38. Lin Y, Zhu M, Su Z. The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemp Clin Trials*. 2015 Nov;45(Pt A):21-5. | CrossRef | PubMed |
39. Burns DK, Chiang C, Welsh-Bohmer KA, Brannan SK, Culp M, O'Neil J, et al. The TOMMORROW study: Design of an Alzheimer's disease delay-of-onset clinical trial. *Alzheimers Dement (N Y)*. 2019 Oct 28;5:661-670. | CrossRef | PubMed |
40. Požgajin I, Požgajin Z, Degmečić D. Placebo and nocebo effect: a mini-review. *Psychiatr Danub*. 2014 Jun;26(2):100-7. | PubMed |
41. Sheldon R, Opie-Moran M. The Placebo Effect in Cardiology: Understanding and Using It. *Can J Cardiol*. 2017 Dec;33(12):1535-1542 | CrossRef | PubMed |
42. Demetriou C, Hu L, Smith TO, Hing CB. Hawthorne effect on surgical studies. *ANZ J Surg*. 2019 Dec;89(12):1567-1576. | CrossRef | PubMed |
43. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ*. 2015 Sep 4;351:h4672. | CrossRef | PubMed |
44. Hodgins GE, Blommel JG, Dunlop BW, Iosifescu D, Mathew SJ, Neylan TC, et al. Placebo Effects Across Self-Report, Clinician Rating, and Objective Performance Tasks Among Women With Post-Traumatic Stress Disorder: Investigation of Placebo Response in a Pharmacological Treatment Study of Post-Traumatic Stress Disorder. *J Clin Psychopharmacol*. 2018 Jun;38(3):200-206. | CrossRef | PubMed |
45. Guyatt G, Rennie D. The principle of intention to treat. In: *Users guide to the medical literature*. Chicago: American Medical Association Press; 2002.
46. Capurro N, Gabrielli N, Letelier L. Importancia de la intención de tratar y el seguimiento en la validez interna de un estudio clínico randomizado. *Rev Méd Chile*. 2004;132(12):1557-60. | CrossRef |
47. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. 2001 Nov 13;165(10):1339-41. | PubMed |
48. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet*. 2002 Mar 2;359(9308):781-5. | PubMed |
49. Ford I, Norrie J. Pragmatic Trials. *N Engl J Med*. 2016 Aug 4;375(5):454-63. | CrossRef | PubMed |
50. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furburg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ*. 2009 May 12;180(10):E47-57. | CrossRef | PubMed |
51. Lesaffre E. Superiority, equivalence, and non-inferiority trials. *Bull NYU Hosp Jt Dis*. 2008;66(2):150-4. | PubMed |
52. Sedgwick P. What is a crossover trial? *BMJ*. 2014 May 9;348:g3191. | CrossRef | PubMed |
53. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. *BMJ*. 2019 Jul 31;366:l4378. | CrossRef | PubMed |

54. Lozano-León J, Ruiz-Peláez J. El experimento clínico. En: *Epidemiología clínica: investigación clínica aplicada*. 2da Ed. Bogotá: Editorial Médica Panamericana; 2015.
55. Mason SA, Rasmussen B, van Loon LJC, Salmon J, Wadley GD. Ascorbic acid supplementation improves postprandial glycaemic control and blood pressure in individuals with type 2 diabetes: Findings of a randomized cross-over trial. *Diabetes Obes Metab*. 2019 Mar;21(3):674-682. | [CrossRef](#) | [PubMed](#) |
56. Baker TB, Smith SS, Bolt DM, Loh WY, Mermelstein R, Fiore MC, et al. Implementing Clinical Research Using Factorial Designs: A Primer. *Behav Ther*. 2017 Jul;48(4):567-580. | [CrossRef](#) | [PubMed](#) |
57. Perman G. Ensayos clínicos por conglomerados (clusters). *Evidencia*. 2017;20(1):22-5. [On line]. | [Link](#) |
58. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. *BMJ*. 2017 Jul 14;358:j3064. | [CrossRef](#) | [PubMed](#) |
59. Bärnighausen T, Rottingen JA, Rockers P, Shemilt I, Tugwell P. Quasi-experimental study designs series-paper 1: introduction: two historical lineages. *J Clin Epidemiol*. 2017 Sep;89:4-11. | [CrossRef](#) | [PubMed](#) |
60. Molina-Arias M, Ochoa-Sangrador C. Ensayo clínico (I). Definición. Tipos. Estudios cuasiexperimentales. *Evid Pediatr*. 2014;10:52.
61. Manterola C, Otzen T. Estudios experimentales 2a parte. Estudios cuasi-experimentales. *Int J Morphol*. 2015;33(1):382-7. | [CrossRef](#) |
62. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*. 2018 May 23;6:e4794. | [CrossRef](#) | [PubMed](#) |
63. Manterola C, Otzen T. Valoración clínica del riesgo, interpretación y utilidad práctica. *Int J Morphol*. 2015;33(3):842-9. | [CrossRef](#) |
64. Molina-Arias M. Cálculo de la reducción del riesgo y el número necesario de pacientes a tratar. *Rev Pediatr Aten Primaria*. 2012;14(56):369-72. | [CrossRef](#) |
65. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Absolute risk reduction, relative risk reduction, and number needed to treat. *Perspect Clin Res*. 2016 Jan-Mar;7(1):51-3. | [CrossRef](#) | [PubMed](#) |
66. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*. 1986;292(6522):746-50. | [CrossRef](#) |
67. Pandis N. Sources of bias in clinical trials. *Am J Orthod Dentofacial Orthop*. 2011 Oct;140(4):595-6. | [CrossRef](#) | [PubMed](#) |
68. Martínez F, Madrid E. [Systematic reviews of biomedical literature: an introduction for busy clinicians]. *Medwave*. 2014 May 23;14(4):e5955. | [CrossRef](#) | [PubMed](#) |
69. Huang YQ, Traore K, Ibrahim B, Sewitch MJ, Nguyen LHP. Reporting quality of randomized controlled trials in otolaryngology: review of adherence to the CONSORT statement. *J Otolaryngol Head Neck Surg*. 2018;47(1):34. | [CrossRef](#) |
70. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010 Aug;63(8):834-40. | [CrossRef](#) | [PubMed](#) |
71. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012 Dec 26;308(24):2594-604. | [CrossRef](#) | [PubMed](#) |
72. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *Pilot Feasibility Stud*. 2016 Oct 21;2:64. | [PubMed](#) |
73. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C, et al. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med*. 2006 Mar 7;144(5):364-7. | [PubMed](#) |
74. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008 Nov 11;337:a2390. | [CrossRef](#) | [PubMed](#) |
75. Montgomery P, Grant S, Mayo-Wilson E, Macdonald G, Michie S, Hopewell S, et al. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension. *Trials*. 2018 Jul 31;19(1):407. | [CrossRef](#) | [PubMed](#) |
76. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *J Clin Epidemiol*. 2016 Aug;76:9-17. | [CrossRef](#) | [PubMed](#) |
77. Des Jarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004 Mar;94(3):361-6. | [PubMed](#) |
78. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996 Feb;17(1):1-12. | [PubMed](#) |
79. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6. Cochrane training; 2019.
80. Navarrete MS. Common statistical misconceptions-plainly explained. *Medwave*. 2019 Jul 2;19(6):e7660. | [CrossRef](#) | [PubMed](#) |
81. Briel M, Bassler D, Wang AT, Guyatt GH, Montori VM. The dangers of stopping a trial too early. *J Bone Joint Surg Am*. 2012 Jul 18;94 Suppl 1:56-60. | [CrossRef](#) | [PubMed](#) |
82. Schou IM, Marschner IC. Meta-analysis of clinical trials with early stopping: an investigation of potential bias. *Stat Med*. 2013 Dec 10;32(28):4859-74. | [CrossRef](#) | [PubMed](#) |
83. Hincapié D, Pérez V, Donado JH. Clasificación de los desenlaces en los ensayos clínicos. *Med UPB*. 2019;38(2):147-57. | [CrossRef](#) |
84. Franco JVA, Arancibia M, Simancas-Racines D, Madrid E. Syntheses of biomedical information: narrative reviews, systematic reviews and emerging formats. *Medwave*. 2018 Nov 27;18(7):e7354. | [CrossRef](#) | [PubMed](#) |

Figura 1. Infografía sobre ensayos clínicos.



Fuente: diseñada por los propios autores.

Correspondencia a
Oficina 1107, Edificio R2
Angamos 655
Reñaca
Viña del Mar
Chile



Esta obra de *Medwave* está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, *Medwave*.