

# Estudios experimentales I: el ensayo clínico aleatorio

Los estudios experimentales son aquellos en los que el equipo investigador controla el factor de estudio, es decir, decide qué sujetos recibirán la intervención que se desea evaluar, así como la forma en que lo harán (dosis, pauta, duración, etc.), de acuerdo con un protocolo de investigación preestablecido. Por tanto, por definición, los estudios experimentales son prospectivos.

El objetivo básico de los estudios experimentales es evaluar los efectos de una intervención, intentando establecer una relación causa-efecto con los resultados observados (son, pues, estudios analíticos). Esta intervención suele ser un tratamiento farmacológico, aunque puede ser cualquier otro tipo de terapéutica, un consejo sanitario, una actividad preventiva, una estrategia diagnóstica o un modelo organizativo, por ejemplo. Lógicamente, dado que la intervención se administra a los sujetos con la finalidad de ser estudiada, son en especial importantes los requisitos éticos de la investigación en seres humanos y solo deben evaluarse intervenciones potencialmente beneficiosas para los sujetos, de las que exista suficiente información previa para justificar la realización del experimento, y diseñando los estudios de acuerdo con los estándares científicos aceptados, tanto éticos como metodológicos.

La gran ventaja de los estudios experimentales sobre los observacionales es que, al controlar el factor de estudio y las condiciones en que se realiza la investigación, disminuye la posibilidad de que otros factores puedan influir sobre los resultados, de manera que proporcionan una mayor confianza en la conclusión obtenida (mayor calidad de la evidencia).

El diseño experimental más importante es el ensayo clínico aleatorio (ECA). Aunque existen diversas formas de diseñarlo, habitualmente esta denominación se utiliza para referirse al diseño paralelo, que es el más conocido y utilizado, y cuyas características generales se describen en este capítulo. Las otras formas de ensayo clínico y del resto de diseños experimentales se abordan en el capítulo siguiente.

## ENSAYO CLÍNICO ALEATORIO

El ECA, en su forma más sencilla y habitual, es un estudio de diseño paralelo con dos grupos para evaluar si un tratamiento es superior al de referencia. Consiste en la selección de una muestra de pacientes y su asignación de forma aleatoria a uno de los dos grupos. Uno de ellos recibe la intervención de estudio, y el otro la de control, que se utiliza como referencia o comparación. En ambos grupos se realiza un seguimiento de forma concurrente durante un período determinado, cuantificando y comparando las respuestas observadas en ambos. La estructura básica se esquematiza en la [figura 5.1](#).

El ECA se considera el mejor diseño disponible para evaluar la eficacia de una intervención sanitaria, ya que es el que proporciona la evidencia de mayor calidad acerca de la existencia de una relación causa-efecto entre dicha intervención y la respuesta observada. Por este motivo, las recomendaciones terapéuticas y las guías de práctica clínica se basan fundamentalmente en las pruebas proporcionadas por este tipo de estudios. Ello ha conducido a que el número de ECA esté aumentando rápidamente y a que cada



### Cuadro 5.1 Ventajas y desventajas de los ensayos clínicos aleatorios

#### Ventajas

- Proporcionan la mejor evidencia de una relación causa-efecto entre la intervención que se evalúa y la respuesta observada.
- Proporcionan un mayor control del factor de estudio.
- La asignación aleatoria tiende a producir una distribución equilibrada de los factores pronósticos que pueden influir en el resultado (potenciales factores de confusión), formando grupos comparables; de este modo, permite aislar el efecto de la intervención del resto de los factores.

#### Desventajas

- Las restricciones éticas impiden que muchas preguntas puedan ser abordadas mediante un ensayo clínico aleatorio.
- Habitualmente se llevan a cabo con participantes muy seleccionados, lo que dificulta la generalización y extrapolación de los resultados.
- A menudo, las intervenciones se administran con pautas rígidas, diferentes de las que se realizan en la práctica habitual, lo que dificulta la generalización y extrapolación de los resultados.
- En general, solo permiten evaluar el efecto de una única intervención.
- Suelen tener un coste elevado, aunque ello depende de la duración del estudio y de la complejidad del protocolo.

pueden considerarse razonablemente alternativas terapéuticas para los pacientes, ya que existen dudas acerca de si una de ellas es superior a la otra. La comparación frente a intervenciones que se sabe que son inferiores, además de ser éticamente inaceptable, conduce a la obtención de resultados favorables a la intervención de estudio, cuya publicación introduce un sesgo en las evidencias disponibles sobre la eficacia de los tratamientos, con las repercusiones que este hecho puede tener sobre las decisiones y las recomendaciones terapéuticas.

**Ejemplo 5.1.** Djulbegovic et al. (2000) evaluaron la calidad de 136 ensayos clínicos publicados sobre el tratamiento del mieloma múltiple en función de la fuente de financiación del estudio. Aunque no observaron diferencias estadísticamente significativas entre la calidad de los ensayos financiados por la industria y los financiados por agencias gubernamentales o sin ánimo de lucro, existían diferencias en relación con el principio de incertidumbre. Mientras que el porcentaje de ensayos financiados por entidades sin ánimo de lucro que favorecían a las nuevas terapias era similar al que favorecía a las de referencia (el 47 frente al 53%), había una clara diferencia en los financiados por la industria a favor de las nuevas terapias (el 74 frente al 26%).

Este hallazgo podía estar relacionado con el hecho de que más ensayos financiados por la industria utilizaban placebo o la ausencia de intervención como grupo de comparación. Los autores concluyen que el sesgo observado en la publicación de la investigación financiada por la industria podría ser la consecuencia de violaciones del principio de incertidumbre.

En términos generales, para seleccionar la intervención de referencia existen tres opciones: placebo, tratamiento activo o ninguna intervención.

Por *placebo* entendemos un preparado que carece de actividad farmacológica, pero cuya apariencia y características organolépticas son idénticas a las del preparado de estudio, lo que permite el enmascaramiento. La finalidad de su uso en investigación es controlar el efecto placebo, que se refiere al efecto psicológico o fisiológico de cualquier medicación, independientemente de su actividad farmacológica, y que depende de factores como la propia personalidad del paciente, las convicciones y el entusiasmo del equipo investigador, las condiciones de administración y las características de la intervención, etc.

Desde el punto de vista de la hipótesis que se pone a prueba, la comparación con

un placebo tiene por objetivo cuantificar el efecto terapéutico del fármaco que se evalúa, ya que la principal ventaja del placebo como alternativa de comparación es controlar los efectos derivados de cualquier característica del tratamiento que no sea el efecto que se está estudiando, incluso los efectos secundarios. Así, si se compara un antihipertensivo con un placebo, se está evaluando si el tratamiento consigue disminuir las cifras de presión arterial más allá de lo que lo haría una sustancia no activa farmacológicamente.

**Ejemplo 5.2.** En el *Lipid Research Clinics* (1984), en el que se comparaba la eficacia de la colestiramina para reducir la morbimortalidad cardiovascular, se observó que algunos efectos secundarios eran bastante frecuentes en el grupo de la colestiramina: se registró estreñimiento en el 39% de los individuos, y en el 27% eructos y flatulencia. Sin embargo, no todo el efecto se debía al fármaco; en el grupo que recibía placebo estos efectos se observaron en el 10 y el 16% de los individuos, respectivamente. Así pues, el uso de placebo permite no solo aislar el efecto terapéutico del fármaco en estudio, sino conocer también la proporción de efectos indeseables que pueden atribuirse a la intervención.

Además, el uso de un placebo permite enmascarar las intervenciones, de forma que los participantes (e incluso los investigadores) desconozcan si reciben tratamiento activo o no.

Cuando existe una opción terapéutica aceptada como eficaz en la situación clínica de interés, esta debería ser la alternativa de comparación. Se ha abusado de la comparación con placebo dado que es más fácil encontrar diferencias estadísticamente significativas que si se compara con otro tratamiento activo. Sin embargo, aunque existe acuerdo universal en que el uso de placebo no es apropiado en situaciones que supongan una amenaza vital y se disponga de alguna intervención eficaz, existe cierta controversia acerca de si puede utilizarse en situaciones en que un retraso en la administración de una intervención eficaz difícilmente suponga un daño permanente en los sujetos.

El uso de otro tratamiento o intervención activos como comparación tiene por objetivo estimar la relación beneficio-riesgo del nuevo tratamiento en una situación clínica concreta. En estos casos, la mejor comparación es el «mejor tratamiento disponible» para dicha situación. Esta no siempre es una elección fácil, ya que en la mayoría de las veces existe un amplio arsenal terapéutico que hace difícil elegir cuál es la mejor alternativa, aunque en algunos casos existen guías o recomendaciones que puedan ser de ayuda en dicha elección.

Además, cuando se emplea un tratamiento activo como referencia, debe prestarse especial atención a la dosis, la pauta y la duración de su administración, para mantener el principio de incertidumbre y no favorecer al nuevo tratamiento al compararlo con otro en inferioridad de condiciones. Se trata de comparar alternativas terapéuticas que, según los conocimientos actuales, podrían estar igualmente indicadas en la situación clínica de estudio.

Cuando se comparan dos tratamientos activos, suele ser conveniente enmascararlos para prevenir posibles sesgos. En muchas ocasiones, para lograrlo es necesario administrar a los sujetos de cada grupo un placebo de la intervención que recibe el otro grupo (*double dummy*). En este caso, el placebo no se utiliza como alternativa de referencia, sino tan solo como una técnica para lograr el enmascaramiento.

**Ejemplo 5.3.** Consideremos un estudio cuyo objetivo sea comparar una monodosis de un antibiótico con una pauta de 10 días para el tratamiento de la infección urinaria. Es conocido que la duración de un tratamiento es un factor que influye sobre su respuesta. Para neutralizar este efecto, el equipo investigador decidió asignar aleatoriamente los individuos a dos grupos, uno de los cuales recibió la monodosis seguida de comprimidos placebo durante 10 días, mientras que el otro recibió en primer lugar una monodosis placebo y, a continuación, el tratamiento hasta completar los 10 días. De este modo, se consiguió mantener el doble ciego y efectuar una comparación lo más imparcial posible.

Algunas veces, por la propia pregunta de investigación, la intervención de referencia más adecuada no es un placebo ni otra intervención específica, sino los cuidados habituales que reciben los pacientes en la consulta (*usual care*). Aunque en teoría es posible comparar el grupo que recibe la intervención de estudio con otro que no recibe ninguna intervención específica, por lo general puede considerarse que, en realidad, el grupo control está recibiendo los cuidados que se suelen prestar para su problema de salud (de otra manera podrían plantearse problemas éticos), por lo que, en realidad, se está comparando con la atención habitual.

**Ejemplo 5.4.** En el *Multiple Risk Factor Intervention Trial* (MRFIT) (1982), individuos de mediana edad con un riesgo elevado de padecer cardiopatía isquémica fueron asignados de forma aleatoria con el fin de evaluar si la intervención simultánea sobre el tabaco, la dieta y la presión arterial prevenía la enfermedad. En el grupo experimental, el control de la hipercolesterolemia y de la hipertensión arterial, así como el consejo dietético y antitabaco, corrían a cargo de un grupo de profesionales altamente especializados, mientras que en el otro grupo no se llevó a cabo ninguna intervención específica, de forma que los sujetos recibían los cuidados habituales dados por sus respectivos médicos de cabecera.

Por otro lado, también debe definirse con claridad la hipótesis que subyace en la comparación. Aunque lo más habitual es que la finalidad sea evaluar si un nuevo tratamiento es más eficaz que el de referencia (estudios de superioridad), también puede diseñarse el estudio para demostrar que no es peor (estudios de no inferioridad) o que ambos son iguales (estudios de equivalencia), ya que la nueva intervención presenta alguna ventaja adicional sobre la estándar, como una mayor facilidad de administración o una mayor seguridad. El ejemplo 5.3 ilustra un estudio de este tipo, puesto que, en caso de que la monodosis no sea inferior, sus ventajas de seguridad y cumplimiento la convertirían en una alternativa preferible a una pauta larga.

## DEFINICIÓN DE LA VARIABLE DE RESPUESTA

La elección de la variable que se utilizará para evaluar la eficacia del tratamiento, cuantificar sus efectos y compararlos con los del grupo de referencia es clave para establecer la relevancia clínica de los resultados que se obtendrán.

Debe seleccionarse la más adecuada, es decir, la que mida los verdaderos resultados de importancia e interés, y no simplemente porque sea fácil de medir o porque se espera que pueda mostrar cambios o diferencias en un corto período de tiempo.

Habitualmente se utilizan variables de importancia clínica, como la supervivencia en el cáncer, las fracturas vertebrales en la osteoporosis, los síntomas o medidas funcionales como la frecuencia de la angina o la tolerancia al ejercicio, o la aparición de sucesos graves e irreversibles, como el infarto agudo de miocardio (IAM) o el accidente cerebrovascular (ACV), por ejemplo. Cada vez se utilizan más las llamadas variables compuestas, que combinan diferentes resultados clínicos en una única variable, lo que permite aumentar el número de desenlaces que se observan. Además de las clínicas, también son relevantes para los pacientes variables como la calidad de vida o la recuperación de una capacidad funcional determinada, por ejemplo.

Algunos estudios utilizan variables subrogadas como sustitutos de un resultado clínicamente relevante. Suelen ser medidas de laboratorio o signos físicos, como la colesteroolemia o las cifras de presión arterial. Sin embargo, existe una importante discusión sobre la adecuación de su uso, ya que en muchas situaciones existen dudas razonables sobre si los cambios producidos en la variable subrogada reflejan de manera adecuada los efectos sobre los verdaderos resultados de interés, por lo que hay que ser especialmente cauteloso en su uso en un ensayo clínico.

Muchos ensayos se centran exclusivamente en los beneficios potenciales de un tratamiento, pero no miden sus efectos sobre otras variables importantes o incluso algunos efectos secundarios y adversos.

## SELECCIÓN DE LA POBLACIÓN

La definición del objetivo hace referencia a la población diana a la que se desea poder extrapolar los resultados. Sin embargo, el ensayo se lleva a cabo sobre una población de estudio definida por unos criterios de selección especificados *a priori*.

Estos criterios identifican sujetos en los que, a la luz de los conocimientos actuales, las intervenciones que se comparan podrían estar igualmente indicadas y, por tanto, potencialmente pueden beneficiarse de ellas. Esto implica que deben excluirse los sujetos en los que una de las alternativas sea preferible a la otra, y aquellos en los que cualquiera de las intervenciones esté contraindicada o podría presentar interacciones. Además de estos criterios generales, pueden establecerse otros criterios de selección para definir una población experimental más adecuada a los objetivos del estudio. La utilización de criterios de inclusión y exclusión estrictos conduce a la obtención de una muestra homogénea, lo que aumenta la validez interna del estudio, pero al ser una muestra restringida limita su capacidad de generalización o extrapolación de resultados. Por otro lado, si se definen criterios muy amplios, la población de estudio será más representativa y las posibilidades de generalizar los resultados serán mayores, pero al ser más heterogénea será más difícil detectar una respuesta al tratamiento y se requerirá un mayor número de individuos.

### Consentimiento informado

Una vez que se ha comprobado que un sujeto cumple todos los criterios de inclusión y ninguno de los de exclusión, antes de incluirlo en el estudio debe dar su consentimiento informado para participar en él.

Previamente a otorgar su consentimiento, un candidato debe recibir información, tanto oral como escrita, sobre lo que supone la experiencia y las posibles consecuencias que pueden derivarse de su participación. Esta información debe incluir como mínimo: *a)* el objetivo del estudio; *b)* las características básicas del diseño (asignación aleatoria, técnicas de enmascaramiento, uso de placebo,

etc.; *c)* los posibles efectos secundarios; *d)* los posibles beneficios; *e)* que el paciente tiene el derecho de abandonar el estudio en cualquier momento, y *f)* que tiene el derecho de formular cuantas preguntas desee relacionadas con la investigación. Todos estos puntos deben explicarse en términos comprensibles para el paciente, evitando el lenguaje académico o científico.

Es muy probable que los sujetos que dan su consentimiento informado difieran en muchos aspectos de los que no lo hacen, incluyendo la motivación y las actitudes hacia la salud y los factores de riesgo de la enfermedad. Aunque este hecho limita la generalización de los resultados, es un imperativo ético que la participación en un ECA debe ser voluntaria y basada en el consentimiento informado.

### Período de preinclusión

Algunos ECA utilizan un período previo a la inclusión de los sujetos y a su asignación a los grupos de estudio (*run-in phase*), durante el cual se realiza el seguimiento de los pacientes con diferentes finalidades.

Por ejemplo, después de obtener su consentimiento, todos los individuos reciben un placebo (o una de las intervenciones, si se desea excluir a los sujetos que presenten efectos secundarios) durante un período especificado (por lo común, unas semanas), tras el cual los sujetos seleccionados se asignan al azar a los grupos para iniciar el estudio.

**Ejemplo 5.5.** En el *American Physicians' Health Study* (Hennekens y Eberlein, 1985), diseñado para evaluar si 325 mg de ácido acetilsalicílico (AAS) administrado a días alternos reduce la mortalidad cardiovascular en pacientes asintomáticos, los 33.223 sujetos que cumplían los criterios de selección establecidos fueron sometidos a un período de preinclusión de 18 semanas durante el cual todos ellos recibieron AAS, de manera que solo fueron incluidos en el estudio los 22.071 sujetos que cumplieron con la intervención durante dicho período y no presentaron efectos secundarios. Estos participantes fueron asignados aleatoriamente a dos grupos: AAS y placebo. Como consecuencia de este proceso de selección tan estricto, a

los 57 meses de seguimiento, prácticamente el 90% de los participantes todavía cumplía con la intervención asignada.

La exclusión de los individuos incumplidores o de los que presentan efectos secundarios aumenta la potencia del estudio y permite una mejor estimación de los efectos globales de la intervención. Sin embargo, aunque aumenta la validez interna del estudio, esta estrategia tiene el inconveniente de que limita la capacidad de extrapolación de los resultados.

En otras ocasiones se utiliza un fármaco activo en el período de preinclusión con la finalidad de usar una respuesta intermedia al tratamiento como criterio de inclusión, por ejemplo, para seleccionar los sujetos que no han respondido al fármaco.

**Ejemplo 5.6.** Un ensayo clínico tenía por objetivo comparar la eficacia de dos alternativas terapéuticas en pacientes que no hubieran respondido a una dosis de 5 mg/día de amlodipino (Marín-Iranzo et al., 2005). Para ello, se diseñó un período de preinclusión durante el cual los pacientes hipertensos recibieron placebo durante 2 semanas y después 5 mg/día de amlodipino durante otras 4 semanas. De los 328 sujetos incluidos inicialmente, 83 fueron excluidos en este período porque se controlaron sus cifras de presión arterial diastólica o presentaron acontecimientos adversos, de manera que finalmente tan solo 245 pacientes fueron asignados aleatoriamente a los dos grupos que se comparaban: un grupo en el que la dosis de amlodipino se aumentó a 10 mg/día y otro grupo al que se administró una combinación de enalapril y nitrendipino. Así pues, la principal utilidad del período de preinclusión en este estudio fue seleccionar a los sujetos no respondedores.

El período de preinclusión también puede utilizarse para seleccionar a los sujetos que cumplen con el tratamiento y las pautas prescritas o para obtener una observación basal más válida y consistente antes de la asignación aleatoria.

Para poder interpretar correctamente los hallazgos de los estudios con período de preinclusión, es importante señalar las diferencias entre las características iniciales

de los sujetos que han sido excluidos durante dicho período y los que son asignados a los grupos de estudio.

## ASIGNACIÓN ALEATORIA

En un ECA, los sujetos incluidos en el estudio se asignan a los grupos siguiendo un método aleatorio. Los beneficios de la asignación aleatoria, que se comentan detalladamente en el capítulo correspondiente, pueden resumirse en dos principales.

En primer lugar, tiende a asegurar la comparabilidad de los grupos, de manera que la comparación entre las intervenciones o tratamientos sea lo más imparcial posible. La asignación aleatoria tiende a producir una distribución equilibrada de las variables entre los grupos, tanto de las conocidas como de las que no lo son. Esta tendencia es tanto mayor cuanto más grande es el tamaño de la muestra. Cuando se estudian pocos pacientes pueden producirse diferencias entre los grupos, a pesar de que los sujetos se hayan asignado de forma aleatoria. El ejemplo más extremo sería cuando solo se estudian dos sujetos; en este caso, asignarlos de forma aleatoria no aporta ningún beneficio, ya que las posibles diferencias entre ambos siguen existiendo. Así pues, es importante tener presente que la asignación aleatoria no garantiza que los grupos sean similares, sino que aumenta la probabilidad de que lo sean. Existen algunas técnicas, como la asignación estratificada o por bloques, que pueden ayudar a garantizar la distribución equilibrada de las variables principales.

En segundo lugar, la asignación aleatoria permite la utilización de técnicas de enmascaramiento, que son muy útiles para obtener una estimación no sesgada de la variable de respuesta.

Dado que la asignación aleatoria es la clave de un ECA, debe realizarse correctamente, de modo que ni las preferencias del médico ni las del paciente influyan en la decisión del grupo al que este es asignado. Por ello, es esencial que se produzca después de que el paciente haya sido incluido en el estudio y que la secuencia de asignación esté oculta para el investigador, de manera que el conocimiento de los grupos no influya en su

decisión de incluirlo en uno u otro. Generalmente, la asignación se realiza tras completar una serie de pruebas (cuestionarios, examen físico, exploraciones complementarias, etc.) para comprobar si los pacientes cumplen o no los criterios de selección. En caso afirmativo, el paciente deberá dar su consentimiento antes de ser incluido. Una vez cumplidos estos dos requisitos se procederá a asignar aleatoriamente a los pacientes, pero no antes. De este modo, la decisión de incluir o no a un sujeto no estará influida por el conocimiento del tratamiento que recibirá.

## TÉCNICAS DE ENMASCARAMIENTO

Las expectativas tanto de los pacientes como de los investigadores pueden influir en la evaluación de la respuesta observada. Este problema se evita utilizando las llamadas *técnicas de ciego* o de *enmascaramiento*, que se definen como aquellos procedimientos realizados con el fin de que algunos de los sujetos relacionados con el estudio (equipo investigador, participantes, etc.) no conozcan algunos hechos u observaciones (básicamente el tratamiento que recibe cada sujeto) que pudieran ejercer un cambio en sus acciones o decisiones y sesgar los resultados. Un estudio que no utiliza técnicas de enmascaramiento se denomina *ensayo abierto*.

### Tipos de enmascaramiento

La técnica del *simple ciego* consiste en que los investigadores, o más frecuentemente los propios participantes, desconozcan qué intervención recibe cada individuo.

Si los investigadores conocen quién recibe cada intervención, o los participantes saben qué tratamiento se les administra, existe la posibilidad de que se examine con mayor minuciosidad cualquier respuesta (aunque sea de modo no intencionado), se pregunte con más detalle por los posibles efectos secundarios o se valore de forma diferente, conscientemente o no, la respuesta observada. Estas preferencias se evitan con la técnica del *doble ciego*, en la que tanto los pacientes como los investigadores des-

conocen el tratamiento administrado. Cuando ambos grupos reciben un tratamiento activo, suele implicar un doble enmascaramiento de las intervenciones, recibiendo cada grupo el tratamiento activo que le corresponde y un placebo del tratamiento del otro grupo.

Por último, existe también la técnica del *triple ciego*, en la que, además, hay otras personas que también desconocen el tratamiento que recibe cada sujeto, ya sea el profesional estadístico que analizará los resultados o bien la persona responsable de decidir si se suspende un tratamiento por la aparición de reacciones adversas o si debe interrumpirse prematuramente el ensayo.

Cuando estos métodos no pueden llevarse a cabo, puede utilizarse la técnica de la evaluación enmascarada de la respuesta, o del *evaluador ciego*. Consiste en que la persona que ha de medir la variable de respuesta desconozca el grupo al que pertenece cada uno de los sujetos, con la finalidad de que la medición se realice e interprete de la misma forma en todos ellos. Esta técnica es muy importante cuando la variable de respuesta es blanda e incluye algún elemento de subjetividad (p. ej., una medida de la intensidad del dolor, un cuestionario sobre síntomas, la percepción de mejoría en relación con una situación previa o la interpretación de una radiografía). En cambio, cuando la variable de respuesta es objetiva (dura), como la mortalidad, esta técnica no es imprescindible.

### Ventajas del enmascaramiento

Un investigador puede tener un cierto prejuicio sobre alguna de las intervenciones. Incluso si es completamente imparcial, los primeros resultados pueden influir sobre sus expectativas y crear un cierto entusiasmo o desilusión, que será difícil de ignorar. Aunque sea de modo inconsciente, cuando en el estudio no se aplica ninguna técnica de enmascaramiento, los errores en la medida de la respuesta siempre favorecen a la intervención preferida, lo que introduce un sesgo en la interpretación de los resultados.

Estas preferencias también se dan en los pacientes. El investigador contagiará su entusiasmo o su pesimismo sobre la nueva

intervención a los participantes en el estudio, lo que influirá sobre la respuesta. Si el paciente está convencido de la eficacia de un nuevo tratamiento, juzgará de forma más benévola sus efectos secundarios y tenderá a sentirse mejor, aunque sea como resultado de la autosugestión. Por ello, siempre que sea posible, es preferible el uso de técnicas de enmascaramiento.

### Limitaciones del enmascaramiento

En los estudios con medicamentos, el enmascaramiento se consigue presentando ambos fármacos con un formato idéntico. Sin embargo, el ciego se puede romper fácilmente si tienen una toxicidad o unos efectos secundarios distintos y bien conocidos, o bien por determinada información que se va recogiendo a lo largo del estudio, por lo que en la práctica el enmascaramiento completo puede ser muy difícil de conseguir.

**Ejemplo 5.7.** En un ECA en el que se comparaban dos fármacos antihipertensivos, los investigadores analizaron la concordancia entre la asignación real, según el método aleatorio, y la que ellos sospechaban en función de los resultados de laboratorio y las reacciones adversas que comunicaba cada paciente. El índice kappa fue de 0,635 (intervalo de confianza del 95% [IC 95%]: 0,372-0,898), lo que indica una concordancia aceptable (García Puig et al., 1995). Este ejemplo muestra cómo el método del doble ciego puede ser desvelado por determinada información. En este caso concreto, la concordancia se basó en la diferente acción de ambos fármacos sobre la excreción renal de albúmina y las diferentes reacciones adversas que comunicaba cada paciente.

En el ejemplo, como comentan los autores del estudio, la mayoría de los efectos adversos aparecieron tarde y los análisis de microalbuminuria se conocieron a los 6 y 12 meses de la asignación aleatoria, por lo que difícilmente pudieron condicionar los resultados.

La aplicación de técnicas de ciego puede ser inviable cuando se evalúan intervenciones no farmacológicas.

**Ejemplo 5.8.** Supongamos un estudio cuyo objetivo es evaluar si el consejo médico

efectuado de forma estructurada es útil en el tratamiento de la depresión posparto, en el que la intervención consiste en varias visitas al domicilio de las pacientes a cargo de profesionales sanitarios previamente entrenados. Las pacientes asignadas al grupo de estudio no podrían estar «ciegas» a la intervención.

El conocimiento, por parte de los individuos, del grupo al que han sido asignados puede conducir a una comparación que no sea imparcial, lo cual es evidente cuando se compara una pauta frente a los cuidados habituales, como en el ejemplo anterior. Aquellos que se encuentran en el primer grupo saben que están recibiendo «algo especial», que puede dar lugar a un evidente y no controlado efecto placebo. Sin embargo, esto es algo inherente a la propia intervención.

Un efecto de signo contrario puede aparecer en el caso de que los individuos sepan que han sido asignados al grupo de no intervención, e intenten compensarlo cambiando su comportamiento y sus actitudes. Este fenómeno recibe el nombre de *contaminación* y suele ocurrir en los estudios sobre hábitos de vida.

**Ejemplo 5.9.** En el *Multiple Risk Factor Intervention Trial (MRFIT) (1982)*, individuos de mediana edad con un riesgo elevado de padecer cardiopatía isquémica fueron asignados de forma aleatoria con el fin de evaluar si la intervención simultánea sobre el tabaco, la dieta y la presión arterial prevenía la enfermedad. En el grupo experimental, el control de la hipercolesterolemia y de la hipertensión arterial, así como los consejos dietético y antitabaco, estaba a cargo de un grupo de profesionales altamente especializados, mientras que en el otro grupo era llevado a cabo por sus respectivos médicos de cabecera. Los investigadores esperaban que el grupo experimental consiguiera reducir sus cifras de colesterol en un 10% si este era mayor de 220 mg/dl, y un 10% la presión diastólica inicial en el caso de que fuera mayor o igual a 95 mmHg, mientras que, en el grupo control, tanto el colesterol como la presión arterial permanecerían estables. También se esperaba que el número de cigarrillos fumados disminuyera en ambos grupos, aunque de manera más pronunciada en el experimental. Sin embargo,

las predicciones no se cumplieron, y en el grupo control los valores de los distintos factores de riesgo también disminuyeron, siempre en la misma dirección y casi con la misma intensidad que en el grupo de intervención. La dilución del efecto hizo que no se encontraran diferencias significativas entre ambos grupos. El tipo de intervención que se evaluaba en el MRFIT hacía imposible la utilización de técnicas de enmascaramiento, por lo que una de las hipótesis podría ser que los sujetos o los médicos, al conocer la intervención, cambiaran su comportamiento en lo referente al ejercicio, el tabaco o la dieta.

En el [cuadro 5.2](#) se enumeran las situaciones en las que no es posible o no es conveniente la utilización del doble ciego.

## SEGUIMIENTO DE LOS SUJETOS

La comparabilidad conseguida con la asignación de los sujetos debe mantenerse a lo largo de todo el estudio. Por ello, la pauta de visitas y exploraciones ha de ser idéntica para todos los participantes.

El seguimiento debe adecuarse a cada problema concreto y ser lo suficientemente prolongado como para permitir que se presente la respuesta esperada. Cuanto menor sea el tiempo de seguimiento, más fácil será mantener el contacto con los participantes, así como el interés y la motivación de los inves-

tigadores por el estudio, y por consiguiente, la probabilidad de pérdidas será menor.

Las fuentes de pérdidas durante el tiempo de observación son diversas. Algunas personas cambiarán de opinión sobre su participación una vez hayan dado su consentimiento y hayan sido asignadas a uno de los grupos, mientras que otras dejarán el estudio a causa de los efectos secundarios de la medicación, por cambios de lugar de residencia o por pérdida de la motivación. El número de pérdidas dependerá de la duración del estudio y de la complejidad del protocolo, por lo que es importante que el seguimiento se haya previsto de forma que evite estos problemas en lo posible. Sin embargo, la posibilidad de que se produzca un sesgo no depende tanto del número de pérdidas como de los motivos y del hecho de que su respuesta a la intervención sea distinta de la observada en los individuos que finalizan el estudio.

**Ejemplo 5.10.** Supongamos un ECA en el que se incluyen 200 sujetos, asignados a dos grupos de 100 individuos cada uno. Finalizan 80 en cada grupo, de los que 40 han presentado un resultado positivo, por lo que la eficacia ha sido del 50% en ambos grupos entre los sujetos que han finalizado el estudio. Sin embargo, este resultado puede estar sesgado por las pérdidas que se han producido. Supongamos que los 20 sujetos que se pierden en el grupo A lo han hecho por motivos no relacionados con el estudio, de forma que 10 de ellos han presentado un resultado positivo. Por tanto, la eficacia en los 100 pacientes del grupo A es de  $(40 + 10) / 100 = 50\%$ . Por el contrario, las pérdidas en el grupo B no se han producido al azar, sino que se han debido a los efectos secundarios de la intervención, de forma que en ninguno de ellos se ha observado un resultado positivo. Por consiguiente, la eficacia en los 100 individuos del grupo B es de  $(40 + 0) / 100 = 40\%$ . Este ejemplo ilustra cómo, aunque se haya producido un mismo número de pérdidas en ambos grupos, si sus motivos son diferentes, pueden introducir un sesgo en los resultados.

Dado que las pérdidas pueden ser una indicación de cómo reaccionan los individuos a las intervenciones, deben ser claramente descritas para poder evaluar su posible impacto sobre los resultados.

### Cuadro 5.2 Situaciones en las que no es posible o no es conveniente la utilización del doble ciego

- Cuando implica riesgos innecesarios para el paciente (p. ej., en el caso de administrar un placebo por vía parenteral repetidamente y durante un tiempo prolongado).
- Cuando no es posible disponer de una formulación galénica adecuada.
- Cuando los efectos farmacológicos permiten identificar fácilmente al menos uno de los fármacos estudiados.
- Cuando, por cualquier circunstancia, se considera que el diseño de doble ciego puede perjudicar la relación entre el médico y el paciente.

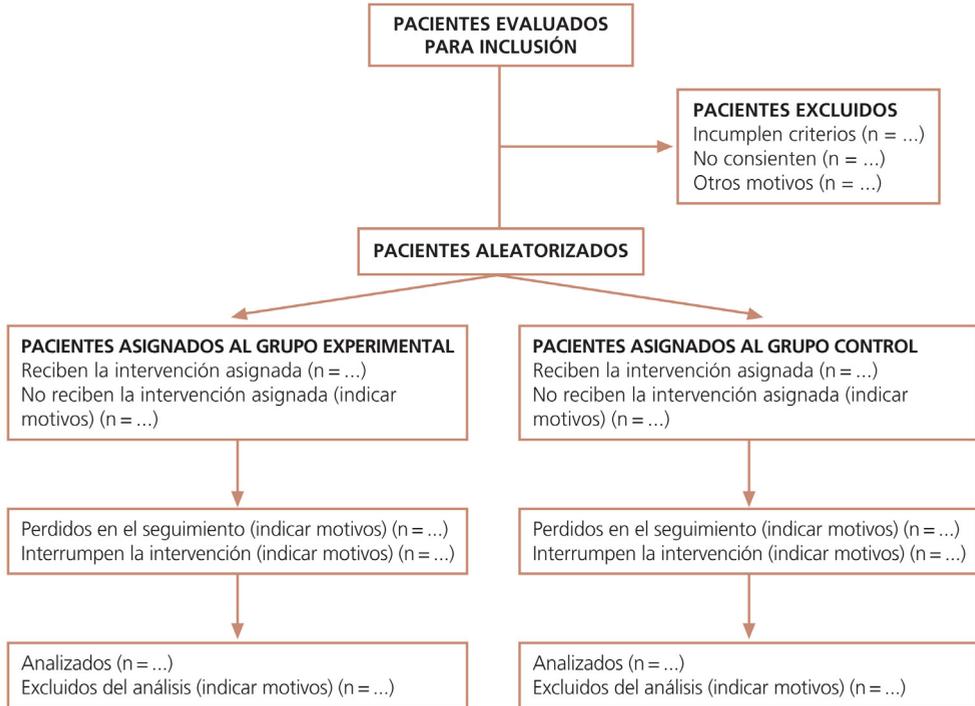


FIGURA 5.2 Diagrama de flujo de pacientes en un ensayo clínico aleatorio de acuerdo con las directrices CONSORT.

La propuesta CONSORT (<http://www.consort-statement.org>), que presenta recomendaciones para mejorar las publicaciones de los ECA, incluye un gráfico que representa el flujo de pacientes a lo largo del estudio y que permite una rápida apreciación de cómo se ha desarrollado (fig. 5.2).

También debe preverse la posibilidad de que determinados sujetos deban ser retirados del estudio por efectos secundarios o por una ausencia de respuesta al tratamiento que reciben. En estos casos, debe estar prevista la intervención que recibirán al ser excluidos del estudio (medicación de rescate).

### ESTRATEGIA DE ANÁLISIS

La estrategia de análisis de un ECA es muy similar a la de cualquier estudio analítico que compara dos o más grupos, y será comentada ampliamente en el capítulo correspondiente. Sin embargo, hay que tener en cuenta algu-

nos aspectos específicos que se esbozan a continuación.

En un ECA pueden presentarse ciertas situaciones que obliguen a considerar si determinados sujetos u observaciones deben ser excluidos del análisis. Según la actitud que se adopte ante estas situaciones, las conclusiones del estudio pueden ser diferentes.

Por un lado, la exclusión de sujetos u observaciones del análisis disminuye la potencia estadística, ya que el número de individuos que se tiene en cuenta es inferior al previsto en el inicio, por lo que existirá una menor capacidad para detectar la diferencia de interés. Sin embargo, lo más importante es que, si el porcentaje de pérdidas o abandonos, o los motivos por los que determinados sujetos no son considerados en el análisis, son diferentes entre los grupos, estas situaciones especiales pueden alterar la comparabilidad. En general, las pérdidas y las violaciones del protocolo, o por lo menos algunas de ellas,

no se producen al azar, sino que pueden depender de los tratamientos administrados. Por tanto, la decisión de excluir determinados sujetos del análisis puede introducir un sesgo y comprometer la validez interna del estudio.

Una opción de análisis es considerar exclusivamente los sujetos de cada uno de los grupos que han cumplido el protocolo del ensayo, han recibido el tratamiento asignado y han finalizado el seguimiento. Esta estrategia se denomina *análisis de casos válidos o por protocolo*.

Si se quiere mantener intacta la comparabilidad conseguida con la asignación aleatoria, la mejor alternativa es evaluar a todos los pacientes según el principio conocido como *análisis por intención de tratar* o según asignación aleatoria, es decir, manteniendo cada sujeto en el grupo al que ha sido asignado, independientemente de si ha cumplido o no con la intervención asignada o de si ha recibido otro tratamiento diferente al asignado. Esta estrategia de análisis se considera la más válida en un ensayo clínico, ya que, además de mantener la comparabilidad de los grupos, es la más conservadora, en el sentido de que es la opción con la que resulta más difícil obtener resultados positivos favorables a la nueva intervención.

Las exclusiones que se producen antes de la asignación aleatoria no originan ningún sesgo en la comparación de los datos, es decir, no afectan a la validez interna. Sin embargo, es importante conocer su número y sus características para poder valorar su impacto sobre la generalización de los resultados.

## FINALIZACIÓN ANTICIPADA

A veces es útil incluir una regla para finalizar el estudio antes de lo previsto, cuando el resultado ya es suficientemente claro. En estas situaciones no es ético que un grupo de sujetos siga recibiendo el tratamiento que se ha mostrado menos eficaz o más perjudicial.

En los estudios con un elevado número de pacientes y un seguimiento de varios años, es recomendable monitorizar los resultados y realizar análisis intermedios en momentos preñados en el protocolo para decidir

si debe finalizarse anticipadamente porque se han alcanzado los objetivos del estudio o por otras causas. La realización de múltiples comparaciones debe tenerse en cuenta en el diseño, tanto al determinar el número de sujetos necesarios como al fijar el grado de significación estadística.

Se recomienda que la decisión de finalizar anticipadamente un ensayo recaiga sobre un grupo de expertos que no esté implicado en el diseño ni en el seguimiento del estudio. Tomar esta decisión como consecuencia de la demostración de efectos beneficiosos debe exigir una evidencia muy convincente. En el [cuadro 5.3](#) se resumen los motivos de interrupción prematura de un ensayo clínico.

**Ejemplo 5.11.** El *American Physicians' Health Study* es un ensayo clínico doble ciego controlado con placebo, diseñado para evaluar si 325 mg de AAS administrado a días alternos reduce la mortalidad cardiovascular en pacientes asintomáticos. Algunas de sus características se han comentado en el ejemplo 5.5. El estudio se suspendió

### Cuadro 5.3 Motivos de interrupción prematura de un ensayo clínico

#### Por datos generados por el ensayo clínico

- Evidencia inequívoca de beneficio o perjuicio del tratamiento.
- Número muy alto e inaceptable de efectos secundarios o colaterales.
- Ausencia de tendencias y de probabilidad razonable de que se demuestren beneficios.

#### Por cuestiones relativas al propio desarrollo del ensayo

- Insuficiente reclutamiento de pacientes en el plazo previsto.
- Mal cumplimiento del tratamiento en un porcentaje elevado de casos.
- Insuficientes recursos financieros.

#### Por datos provenientes de fuentes externas al ensayo

- Datos de otros ensayos que proporcionen una evidencia inequívoca de beneficio o perjuicio del tratamiento.
- Nuevos desarrollos que dejen obsoleto el seguimiento terapéutico objeto del ensayo.

de forma anticipada tras un promedio de 60,2 meses de seguimiento de los 22.071 participantes, al observarse una importante reducción relativa del riesgo de IAM del 44% en el grupo de AAS en relación con el de placebo (riesgo relativo = 0,56; IC 95%: 0,45-0,70;  $p < 0,00001$ ). Se consideró que la evidencia de que el AAS reducía el riesgo de IAM era suficientemente concluyente para justificar la finalización anticipada del ensayo, a pesar de que las pruebas sobre la incidencia de ACV y la mortalidad cardiovascular total eran todavía insuficientes, dado el escaso número de eventos observados (Steering Committee of the Physicians' Health Study Research Group, 1989).

## ESTUDIOS DE EQUIVALENCIA Y NO INFERIORIDAD

En ocasiones, un ECA no se diseña con el objetivo de determinar la superioridad de la intervención de estudio en relación con la de comparación, sino de comprobar si ambas son iguales (estudios de equivalencia) o si, al menos, la primera no es inferior a la segunda (estudios de no inferioridad). Estos estudios se realizan cada vez con más frecuencia, y tienen interés fundamentalmente cuando una nueva intervención presenta ventajas por su seguridad, comodidad de administración o coste, por ejemplo, o simplemente puede representar una nueva alternativa terapéutica, por lo que no es imprescindible que muestren una mayor eficacia.

Deben diferenciarse de los llamados estudios de bioequivalencia, que corresponden a ensayos de fase I que comparan dos formulaciones o métodos de administración de un fármaco, con la intención de demostrar que son intercambiables. Las variables de respuesta que se utilizan son medidas farmacocinéticas. Suelen realizarse con un número reducido de sujetos y utilizando diseños cruzados.

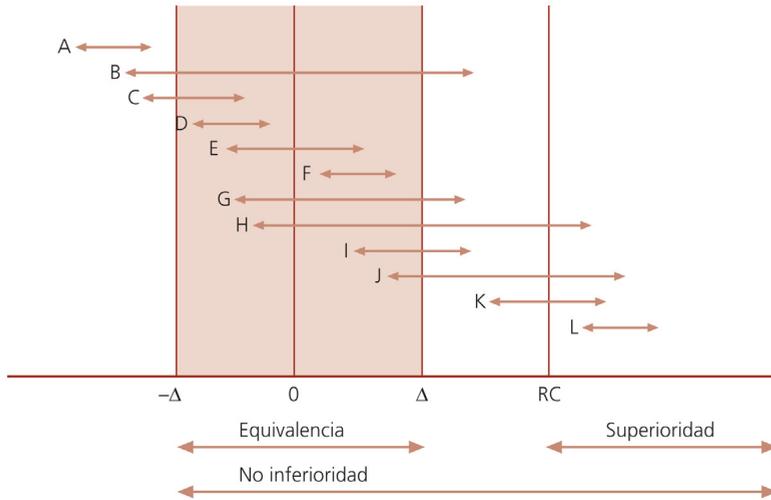
El diseño de los estudios de equivalencia y no inferioridad presenta algunas diferencias importantes en relación con los de superioridad, y debe ser especialmente riguroso. En primer lugar, debe asegurarse la imparcialidad de la comparación, es decir, que las condiciones del ensayo no favorecen

a ninguna de las intervenciones por encima de la otra (dosis y pautas óptimas para ambas, evitar subgrupos de pacientes que podrían manifestar de forma distinta los efectos de los tratamientos, variables y tiempo de valoración adecuados). Dado que el hecho de que dos intervenciones sean equivalentes no implica necesariamente que ambas sean eficaces, es importante que el ensayo se diseñe con criterios similares a los de los estudios en los que la intervención de comparación ya ha demostrado suficientemente su eficacia.

La equivalencia absoluta no puede demostrarse por completo. Cuando un estudio comparativo no detecta ninguna diferencia entre dos tratamientos, puede deberse tanto a que ambos tengan una eficacia similar como a la incapacidad del estudio para detectar una diferencia entre ellos. Por ello, es fundamental que los estudios de equivalencia se diseñen de manera que tengan la sensibilidad suficiente para detectar una diferencia relevante entre los tratamientos que se comparan.

Un aspecto clave es establecer el margen de equivalencia, que corresponde a la máxima diferencia entre tratamientos que se considera clínicamente aceptable por su escasa relevancia clínica. Este valor se conoce como delta ( $\Delta$ ), y la finalidad del estudio es precisamente descartar estadísticamente su existencia. Se trata de demostrar que la nueva intervención es lo suficientemente similar a la de referencia como para considerarlas equivalentes desde el punto de vista clínico.

**Ejemplo 5.12.** Birtwhistle et al. (2004), ante la escasa evidencia sobre la periodicidad adecuada de las visitas de seguimiento de los pacientes hipertensos, compararon el control de la presión arterial, la satisfacción y el cumplimiento del tratamiento por parte de estos pacientes, según si las visitas se realizaban cada 3 o cada 6 meses. Para ello, diseñaron un ensayo clínico en el que participaron 50 médicos de familia de Canadá. Los 609 pacientes incluidos fueron asignados a dos grupos, y de cada uno se realizó el seguimiento con una de las estrategias que se comparaban. La asignación fue aleatoria, estratificada por médico y por bloques de ocho pacientes para asegurar que cada médico tenía un



$\Delta$ : margen de equivalencia.

RC: mínima diferencia de relevancia clínica.

*Hipótesis de superioridad*: si la conclusión se basara exclusivamente en la significación estadística, cuatro estudios se considerarían «negativos» (B, E, G y H), cinco positivos a favor del tratamiento (F, I, J, K y L) y tres en contra (A, C y D). Si se ha fijado a priori la mínima diferencia de relevancia clínica (RC) para considerar que el nuevo tratamiento es superior al estándar, solamente el ensayo L sería concluyente a favor del mismo, mientras que los ensayos H, J y K no lo serían totalmente.

*Hipótesis de equivalencia*: solamente tres estudios (D, E y F) podrían concluir la equivalencia de ambas intervenciones, ya que todo el IC queda incluido dentro del área definida por los límites  $\pm$  delta. Los ensayos A, K y L demostrarían la no equivalencia de las intervenciones (en A por la inferioridad del nuevo tratamiento, y en K y L por su superioridad), mientras que en el resto de los ensayos los resultados no serían concluyentes.

*Hipótesis de no inferioridad*: todos los ensayos podrían concluir no inferioridad excepto el ensayo A, que concluiría inferioridad, y los ensayos B y C, que no podrían descartarla.

**FIGURA 5.3** Interpretación de los resultados de un ensayo según la hipótesis del estudio.

número de pacientes similar en cada grupo. El diseño fue planteado como un estudio de equivalencia, con la hipótesis de que la verdadera diferencia en el porcentaje de pacientes controlados entre los grupos sería inferior al 10%. A los 3 años de seguimiento, el porcentaje de pacientes controlados, el grado de satisfacción y el cumplimiento del tratamiento fueron similares en ambos grupos, sin diferencias relevantes ni estadísticamente significativas. Los autores concluyeron que la estrategia de seguimiento de visitas cada 6 meses es suficiente, ya que es equivalente a la de cada 3 meses.

El análisis de los resultados de estos estudios se basa en el cálculo del intervalo de confianza de la diferencia entre las respuestas observadas (fig. 5.3). En los ensayos de equivalencia, si todo el intervalo cae dentro del rango de equivalencia definido por el

valor delta prefijado, puede concluirse que las intervenciones son equivalentes, mientras que si cae fuera de dicho rango, se concluye que no son equivalentes. En el resto de situaciones, los resultados del estudio no son concluyentes. En los ensayos de no inferioridad, el enfoque es similar, pero se tiene en cuenta tan solo el margen de no inferioridad definido por el valor delta.

En los estudios de equivalencia y no inferioridad, la estrategia de análisis por intención de tratar no es la más conservadora, ya que tiende a hacer más similares los resultados de ambos grupos. El análisis por protocolo tiende precisamente a destacar cualquier diferencia entre los grupos más que a disminuirla, por lo que es preferible en este tipo de estudios realizar ambas estrategias de análisis y, si son coincidentes, los resultados serán más concluyentes.

## ENSAYOS PRAGMÁTICOS Y ENSAYOS EXPLICATIVOS

Al diseñar un ECA deben tomarse numerosas decisiones sobre el comparador, los criterios de selección, la pauta de seguimiento, etc. En general, en el diseño pueden adoptarse dos posturas contrapuestas: explicativa y pragmática (tabla 5.1).

La *actitud explicativa* consiste en establecer criterios de selección muy estrictos, que definan una población muy homogénea, con escasa variabilidad, buena cumplidora, en la que sea más fácil obtener datos de calidad y en la que exista una mayor probabilidad de encontrar un efecto, si existe. Esta actitud tiene algunos inconvenientes, ya que si los criterios son muy estrictos puede no encontrarse un número suficiente de sujetos o no detectarse un efecto en un determinado

subgrupo por no haberse considerado antes. Además, se trata de una población altamente seleccionada, lo que dificulta la generalización o extrapolación de los resultados. Precisamente esta es una de las principales limitaciones de los ECA, tal como se realizan habitualmente, como base para la elaboración de recomendaciones terapéuticas.

La *actitud pragmática* consiste en establecer unos criterios de selección amplios, que definan una población heterogénea, más representativa de la población diana. Su inconveniente es que se pierde cierto grado de control sobre la situación y puede diluir o enmascarar una asociación o un efecto existentes, por lo que se complica la interpretación de los resultados.

La elección de una u otra actitud dependerá de la propia actitud del investigador y del objetivo concreto que se desee alcanzar.

**TABLA 5.1** Actitud explicativa y actitud pragmática en el diseño de un ensayo clínico aleatorio

	Actitud explicativa	Actitud pragmática
Finalidad	Aumentar el conocimiento científico	Ayudar en la toma de decisiones
Objetivo	Eficacia	Efectividad
Problema de salud	Más adecuada para los agudos	Más adecuada para los crónicos
Condiciones de realización	Experimental, «de laboratorio»	De consulta, de práctica clínica
Intervención de estudio	Simple y muy estandarizada Dosis fijas	A menudo compleja Dosis a menudo variables
Intervención de referencia	Habitualmente placebo	A menudo cuidados habituales u otra intervención activa
Respuesta	Acción farmacológica	Efecto terapéutico
Tratamientos concomitantes	Se evitan	A menudo se permiten
Enmascaramiento	Doble ciego	A menudo abierto
Muestra de sujetos	Homogénea Cumplidora A menudo, tamaño menor	Heterogénea Representativa Tamaño mayor
Criterios de selección	Restringidos	Amplios
Período de preinclusión	Frecuente	No

© Elsevier. Fotocopiar sin autorización es un delito.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Birtwhistle RV, Godwin MS, Delva MD, Casson RI, Lam M, MacDonald SE, et al. Randomised equivalence trial comparing three month and six month follow up of patients with hypertension by family practitioners. *BMJ* 2004;328:204.
- Djulfbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635-8.
- García Puig J, Mateos Antón F, Gil Aguado A, Barcina Sánchez C. ¿Puede conocerse el «doble ciego» antes de desvelar los códigos de identificación? *Med Clin (Barc)* 1995;105:437-8.
- Hennekens CH, Eberlein K. A randomized trial of aspirin and beta-carotene among US physicians. *Prev Med* 1985;14:165-8.
- Lipid Research Clinics Program The lipid research clinics coronary primary prevention trial results. 1. Reduction in incidence of coronary heart disease. *JAMA* 1984;251:351-64.
- Marín-Iranzo R, de la Sierra-Iserte A, Roca-Cusachs A, Oliván-Martínez J, Redón-Mas J, Luque-Otero M, et al. Estudio doble ciego de la eficacia y la seguridad de la combinación a dosis fija de 10 mg de enalapril/20 mg de nifedipino en comparación con el incremento de dosis de amlodipino en pacientes con hipertensión esencial no controlada con 5 mg de amlodipino. *Rev Clin Esp* 2005;205:418-24.
- Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial. *JAMA* 1982;248:1465-77.
- Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989;321:129-35.

## BIBLIOGRAFÍA

- Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc)* 2002;118:701-3.
- Boissel JP. Planning of clinical trials. *J Intern Med* 2004;255:427-38.
- Chow SC, Shao J. A note on statistical methods for assessing therapeutic equivalence. *Control Clin Trials* 2002;23:515-20.
- Dasgupta A, Lawson KA, Wilson JP. Evaluating equivalence and noninferiority trials. *Am J Health Syst Pharm* 2010;67:1337-43.
- Djulfbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *JAMA* 2001;285:1206-8.
- Elie C, De Rycke Y, Jais JP, Marion-Gallois R, Landais P. Methodological and statistical aspects of equivalence and non inferiority trials. *Rev Epidemiol Sante Publique* 2008;56:267-77.
- Ellenberg SS, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: practical issues and specific cases. *Ann Intern Med* 2000;133:464-70.
- Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat

- principle and excluding patients from analysis. *BMJ* 2002;325:652-4.
- Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. *Fundamentals of clinical trials*. 5th ed. New York: Springer; 2015.
- Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003;3:28.
- Gomberg-Maitland M, Frison L, Halperin JL. Active-control clinical trials to establish equivalence or noninferiority: methodological and statistical concepts linked to quality. *Am Heart J* 2003;146:398-403.
- Green SB. Design of randomized trials. *Epidemiol Rev* 2002;24:4-11.
- Jiménez Villa J, Argimón Pallas JM. Ensayos clínicos aleatorizados. En: Jiménez Villa J, Argimón Pallas JM, Martín Zurro A, Vilardell Tarrés M, editores. *Publicación científica biomédica: cómo escribir y publicar un artículo de investigación*. 2.ª ed. Barcelona: Elsevier; 2016. p. 139-62.
- Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36-9.
- Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:167-89.
- McMahon AD. Study control, violators, inclusion criteria and defining explanatory and pragmatic trials. *Stat Med* 2002;21:1365-76.
- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
- Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ* 2001;165:1339-41.
- Pablos Méndez A, Barr G, Shea S. Run-in periods in randomized trials: implications for the analysis of results in clinical practice. *JAMA* 1998;279:222-5.
- Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. *Epidemiol Rev* 2002;24:26-38.
- Pocock SJ. When to stop a clinical trial. *BMJ* 1992;305:235-40.
- Pocock SJ, Clayton TC, Stone GW. Design of major randomized trials: part 3 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol* 2015;66:2757-66.
- Roland M, Torgerson BJ. What are pragmatic trials? *BMJ* 1998;316:285.
- Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. *JAMA* 2002;288:2406-7.
- 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;273:408-12.
- Scott IA. Non-inferiority trials: determining whether alternative treatments are good enough. *Med J Aust* 2009;190:326-30.
- Stanley K. Design of randomized controlled trials. *Circulation* 2007;115:1164-9.
- Stanley K. Evaluation of randomized controlled trials. *Circulation* 2007;115:1819-22.