

Ensayos clínicos aleatorizados (CONSORT)

Albert Cobos-Carbó

Unidad de Bioestadística. Departamento de Salud Pública. Facultad de Medicina.
Universidad de Barcelona. Barcelona. España.

Se ha puesto de manifiesto repetidamente que la información facilitada en las publicaciones de ensayos clínicos es muchas veces insuficiente o inexacta y que ciertos problemas metodológicos se asocian a estimaciones exageradas del efecto de las intervenciones sanitarias. Para mejorar la calidad de la comunicación de los ensayos clínicos, un grupo de científicos y editores desarrolló el *CONSORT statement (Consolidated Standards of Reporting Trials)*, una guía de 22 puntos (y un diagrama de flujo), que pueden utilizar autores, editores, revisores y lectores. Tras su publicación en 1996, diversas revistas y grupos editoriales adoptaron la guía CONSORT. En 1999 se elaboró una segunda versión revisada que se publicó en 2001. En este artículo se presenta la traducción al castellano de los dos elementos que integran la guía CONSORT revisada, el diagrama de flujo y la lista-guía de 22 puntos, acompañada de un breve comentario sobre cada uno de ellos. Las publicaciones previas del *CONSORT statement* y otros recursos útiles, como ejemplos de lo que se consideran comunicaciones correctas, pueden obtenerse en la página web de CONSORT (<http://www.consort-statement.org>).

Palabras clave: Publicación/*estándares. Control de calidad. Ensayos clínicos controlados. Guías.

Randomized clinical trials (CONSORT)

It has been repeatedly shown that the information supplied in publications of clinical trials is frequently insufficient or inaccurate and that some methodologic problems are associated with exaggerated estimates of the effect of healthcare interventions.

To improve the quality of reports of clinical trials, a group of scientists and editors developed the *CONSORT statement (Consolidated Standards of Reporting Trials)*, a 22-item checklist (plus flow diagram), that can be used by authors, editors, reviewers, and readers. After publication in 1996, CONSORT was adopted by several journals and editorial groups. In 1999, a second version was drawn up, which was published in 2001.

This article presents the Spanish translation of the two elements integrating the revised CONSORT, the flow diagram and the 22-item checklist, and provides a short comment on each of them. Previous publications of the *CONSORT statement* and other useful resources such as examples of what are considered good communications may be obtained from the CONSORT web site (<http://www.consort-statement.org>).

Key words: Publishing/standards. Quality assurance. Randomized controlled trials/standards. Guidelines.

Introducción

Los ensayos clínicos aleatorizados (ECA) bien diseñados y correctamente ejecutados proporcionan la mejor evidencia sobre el efecto de las intervenciones sanitarias, pero una metodología inadecuada puede resultar en estimaciones exageradas de este efecto^{1,2}.

Para poder evaluar la metodología de un ECA a partir de su publicación, es preciso que su diseño, ejecución y análisis se describan detalladamente, con exactitud y con transpa-

rencia. Sin embargo, se ha constatado repetidamente que la información facilitada en las publicaciones es muchas veces insuficiente³⁻⁶, o inexacta (por abuso de términos como análisis por intención de tratar⁷, o aleatorización⁸), y que ciertos problemas metodológicos son frecuentes⁹⁻¹¹.

Con ánimo de mejorar la situación, a principios de los años noventa, dos grupos de editores, profesionales del ensayo clínico y metodólogos, elaboraron separadamente recomendaciones para comunicar ensayos clínicos^{12,13}. Posteriormente, ambos grupos reunieron sus esfuerzos y consensuaron una propuesta conjunta que se publicó con el título de *CONSORT statement*, acrónimo de *Consolidated Standards of Reporting Trials*¹⁴. CONSORT es básicamente una lista-guía de los puntos que se consideran críticos y que por tanto deberían incluirse en todo informe o comunicación de un ensayo clínico, y un diagrama para ilustrar el flujo de individuos participantes durante el ensayo. En principio, la guía CONSORT se desarrolló para ensayos clínicos de grupos paralelos y, aunque gran parte de su contenido es también válido para otros diseños, se trabaja en extensiones o adaptaciones para otro tipo de ensayos^{15,16}.

El objetivo de la guía CONSORT es facilitar la lectura crítica y la interpretación de los ECA, orientando a los autores, revisores y editores acerca de cómo debe comunicarse este tipo de estudios. Desde su publicación en 1996, diversas revistas y grupos editoriales (incluido el grupo de Vancouver) han adoptado la guía CONSORT, y posteriormente se ha verificado un aumento en la calidad de la comunicación de los ECA^{17,18}. Sin embargo, también se puso de manifiesto ciertas limitaciones¹⁹, por lo que en 1999 se emprendió la revisión de la versión inicial, lo que dio lugar a una versión revisada que se publicó en 2001²⁰⁻²².

En este artículo se presenta la traducción de los 22 puntos (tabla 1) y del diagrama de flujo (fig. 1) que constituyen la guía CONSORT, acompañados de un breve comentario. En una publicación previa²³ y también en internet (<http://www.consort-statement.org>) se puede consultar la justificación de los 22 puntos incluidos en la guía, así como ejemplos publicados de lo que se consideran comunicaciones correctas.

Puntos de la lista de comprobación

Título y resumen

1. *Título y resumen.* Para facilitar su indexación como ECA en las bases de datos bibliográficas y su identificación en las búsquedas²⁴, se recomienda mencionar el método de asignación de los individuos a los grupos de tratamiento tanto en el resumen como en el título del artículo.

Introducción

2. *Antecedentes.* Resumir y referenciar el conocimiento previo, así como razonar la necesidad del ensayo, son condiciones para cumplir con el principio ético de evitar la exposición innecesaria de los individuos a los riesgos de una investigación poco justificada²⁵. Algunos ensayos clínicos

Correspondencia: Dr. A. Cobos-Carbó.
Unidad de Bioestadística. Departamento de Salud Pública.
Facultad de Medicina. Universidad de Barcelona.
Casanova, 143. 08036 Barcelona. España.
Correo electrónico: acobos@ub.edu

TABLA 1

Lista de comprobación para publicar un ensayo clínico aleatorizado

	Punto n.º	Descriptor	Informado en página número
Título y resumen	1	Cómo se asignaron los pacientes a las intervenciones (p. ej., "asignación aleatoria" o "aleatorizado")	
Introducción			
Antecedentes	2	Antecedentes científicos y razón de ser del estudio	
Métodos			
Participantes	3	Criterios de selección de pacientes y ámbito y lugares en que se recogieron los datos	
Intervenciones	4	Detalles precisos de las intervenciones pretendidas para cada grupo y cómo y cuándo efectivamente se administraron	
Objetivos	5	Objetivos e hipótesis específicos	
Resultados	6	Definición clara de las medidas de los resultados principal y secundarios y, cuando proceda, de cualquier método utilizado para mejorar la calidad de las medidas (p. ej., observaciones múltiples, entrenamiento de evaluadores)	
Tamaño muestral	7	Cómo se determinó el tamaño muestral y, cuando proceda, explicación de cualquier análisis intermedio y reglas de interrupción del ensayo	
Aleatorización			
Generación de la secuencia	8	Método utilizado para generar la secuencia de asignación aleatoria, incluido el detalle sobre cualquier restricción (p. ej., bloques, estratificación)	
Asignación oculta	9	Método utilizado para implementar la secuencia de asignación aleatoria (p. ej., contenedores de medicación numerados, aleatorización centralizada por teléfono), y clarificar si la secuencia se mantuvo oculta hasta el momento de la asignación	
Implementación	10	Quién generó la secuencia de asignación, quién incluyó a los participantes y quién los asignó a los grupos	
Cegado (enmascaramiento)	11	Si los participantes, quienes administraron las intervenciones o quienes evaluaron los resultados, conocían o no la intervención asignada. Si procede, cómo se evaluó el éxito del enmascaramiento	
Métodos estadísticos	12	Métodos estadísticos empleados para comparar los grupos en el resultado principal; métodos utilizados en análisis adicionales, como análisis de subgrupos o análisis ajustados	
Resultados			
Flujo de participantes	13	Flujo de participantes en cada fase (se recomienda un diagrama). Específicamente, para cada grupo, documentar los números de participantes asignados aleatoriamente, que recibieron el tratamiento pretendido, que completaron el protocolo del estudio y a los que se incluyó en el análisis del resultado principal. Describir las desviaciones del protocolo planificado, y los motivos	
Reclutamiento	14	Fechas que limitan los períodos de reclutamiento y de seguimiento	
Datos basales	15	Características demográficas y clínicas basales en cada grupo	
Números analizados	16	Número de participantes (denominador) de cada grupo incluidos en cada análisis, y si el análisis se realizó «por intención de tratar». Expresar los resultados mediante números absolutos cuando sea factible (p. ej., 10/20 en lugar de 50%)	
Resultados y estimación	17	Para cada resultado principal y secundario, un resumen de resultados por grupo y el efecto estimado y su precisión (p. ej., IC del 95%)	
Análisis complementarios	18	Considerar la multiplicidad, e informar sobre cualquier otro análisis realizado, incluidos análisis de subgrupos y análisis ajustados, indicando los preespecificados y los exploratorios	
Eventos adversos	19	Todos los eventos adversos o efectos colaterales importantes en cada grupo de intervención	
Discusión			
Interpretación	20	Interpretación de los resultados, teniendo en cuenta las hipótesis del estudio, las fuentes de sesgo o imprecisión potenciales y los peligros asociados a la multiplicidad de análisis y de variables	
Generalización	21	Generalización (validez externa) de los hallazgos del ensayo	
Evidencia global	22	Interpretación general de los resultados en el contexto de la evidencia actual	

podrían haberse evitado haciendo una revisión sistemática de los previos²⁶. Por ello, la introducción debe hacer referencia a los ensayos clínicos previos similares y a posibles revisiones sistemáticas de dichos ensayos, o constatar su ausencia.

Métodos

3. Participantes. En los ensayos clínicos, la población de individuos definida por la condición clínica de interés suele restringirse mediante criterios de selección tales como la edad, la ausencia de comorbilidad o de comedición. Todos los criterios de selección (la distinción entre criterios de inclusión y de exclusión es innecesaria²⁷) utilizados deben

definirse explícitamente, así como los métodos utilizados para reclutar a los pacientes. Además de los criterios de selección, debe especificarse el ámbito y los centros en que se ha realizado el ensayo, porque pueden ser determinantes de la validez externa (posibilidad de generalización) de los resultados.

4. Intervenciones. Debe describirse las características de todas las intervenciones, incluidas las utilizadas como control (p. ej., placebo o práctica clínica habitual), particularmente las características de las que, presumiblemente, pueda depender el efecto. Dada la diversidad de posibles intervenciones (fármacos, cirugía, programas educativos, etc.), es difícil estandarizar la descripción de las características relevantes.

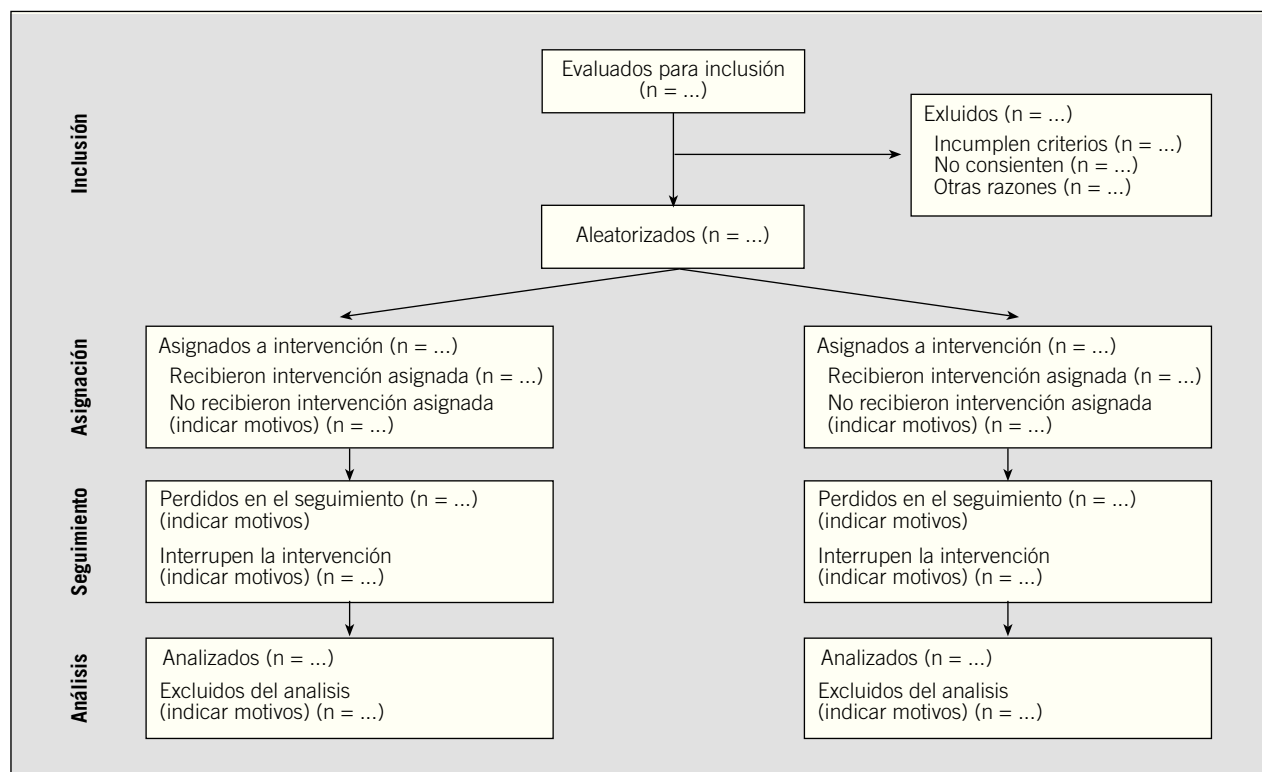


Fig. 1. Diagrama de flujo de participantes.

En algunos casos puede ser imprescindible describir el entrenamiento o la experiencia de quienes practicaron la intervención, además de la técnica empleada (p. ej., intervenciones quirúrgicas)²⁸. Si se trata de intervenciones complejas, como combinaciones de tratamientos, hay que describir los distintos componentes.

5. Objetivos. Debe mencionarse explícitamente los objetivos (o las correspondientes hipótesis) específicos del ensayo y, en el caso de los ensayos confirmatorios, es necesario distinguir el principal de los secundarios.

6. Resultados. El término *resultados* se refiere aquí a las variables utilizadas para evaluar comparativamente las intervenciones. Es necesario definir las variables utilizadas y distinguir la *principal* de las *secundarias*. La variable principal debe ser la que se consideró más importante al diseñar el ensayo y se utilizó para determinar el tamaño muestral. Es fundamental que su definición sea inequívoca: cuando la misma variable se evalúa repetidamente (distintos tiempos o métodos), hay que aclarar cuál de ellas, o qué combinación de ellas, es la que se consideró principal. En general, es recomendable definir una única variable principal para evitar problemas de multiplicidad (véase el punto 18); pero si hay más de una, es necesario tenerlos en cuenta y discutirlos. Si los hay, debería emplearse instrumentos (incluyendo escalas²⁹, p. ej. de calidad de vida) y procedimientos de medida descritos previamente³⁰, e idealmente consensuados, estandarizados o, al menos, de propiedades conocidas, describiendo estas propiedades³¹.

7. Tamaño muestral. Cuando se diseña un ensayo clínico, es ética y científicamente importante asegurar que el análisis principal tendrá una potencia adecuada, es decir, una probabilidad razonablemente alta de demostrar diferencias clínicamente relevantes si éstas existen. Por ello, es importante des-

cribir cómo se determinó el tamaño muestral, incluidos las estimaciones utilizadas en los cálculos y los posibles ajustes para compensar pérdidas, y explicar cualquier discrepancia entre el tamaño muestral pretendido y el finalmente conseguido. No es particularmente interesante calcular *a posteriori* la potencia de un ensayo porque los intervalos de confianza (véase el punto 17) indican la imprecisión de las estimaciones³².

En algunos estudios se realizan análisis intermedios para decidir si vale la pena continuar con la inclusión de pacientes o, por el contrario, interrumpir el ensayo. Estos estudios requieren métodos de análisis que tengan en cuenta la multiplicidad de análisis³³⁻³⁵, y es necesario describir los que se han utilizado y el número de análisis realizados.

8. Generación de la secuencia aleatoria. En general, la asignación aleatoria se considera uno de los pilares fundamentales en que descansa la credibilidad de los resultados de un ensayo clínico (algunos autores preconizan un procedimiento de asignación, no necesariamente aleatorio, denominado minimización³⁶). El término *aleatorio* tiene un significado técnico preciso que no siempre es bien comprendido³⁷ y que a veces se utiliza con ligereza⁸. En ocasiones la descripción de los métodos de asignación utilizados no permite juzgar si esta asignación fue realmente aleatoria³⁸. Por ello no basta con afirmar que la asignación fue aleatoria, sino que debe explicarse cómo se generó la secuencia de asignación (p. ej., mediante tablas de números aleatorios o mediante funciones generadoras de números aleatorios de programas informáticos), e indicar la razón de asignación (p. ej. 3:1) y las restricciones impuestas en el esquema de aleatorización (p. ej., ninguna, bloques de tamaño constante o variable, estratificación). Si se utilizó minimización, es necesario explicar cuáles fueron los factores de minimización y si se incluyó un componente aleatorio o fue enteramente determinista.

9. *Asignación oculta*. Para evitar que la asignación de los tratamientos sea sesgada no basta con generar una secuencia de asignación de tratamientos aleatoria³⁹. La forma en que se utilice esta secuencia es crítica. Si quien ha de decidir la inclusión o no de un individuo en el ensayo conoce la secuencia de asignación de tratamientos, la decisión de inclusión puede estar condicionada (consciente o inconscientemente) por el tratamiento que recibirá el paciente en caso de que se lo incluya. El procedimiento de asignación debe garantizar que el tratamiento asignado es impredecible para quien decide la inclusión de individuos⁴⁰, lo que se consigue manteniendo oculta la secuencia de asignación (*concealed allocation*)¹. Para ello se puede utilizar distintos procedimientos, como la aleatorización telefónica por terceros o sistemas de asignación automatizados⁴¹. Dada la importancia del proceso de asignación de los tratamientos, es necesario describir el método utilizado.

10. *Implementación*. Para que se pueda juzgar si la asignación pudo ser predecible o no, además de describir el método utilizado, es particularmente importante explicar quién preparó la secuencia de asignación, quién decidió la inclusión de los individuos y quién realizó la asignación de los tratamientos.

11. *Cegado (enmascaramiento)*. En un ensayo clínico, el término *cegado*, o *a ciegas*, se refiere a que los participantes ignoran qué tratamiento se ha asignado a cada individuo. Para conseguirlo, los distintos tratamientos deben *enmascarse* de forma que sean indistinguibles. Los ensayos sin enmascaramiento de los tratamientos se describen como ensayos *abiertos*.

El desconocimiento del tratamiento asignado por parte de quienes lo reciben, de quienes lo administran, de quienes evalúan el resultado⁴² y de quienes analizan los datos⁴³ imposibilita que el prejuicio influya en el resultado del ensayo. El término *doble cegado*, muy utilizado en las publicaciones, es un tanto ambiguo: implica que quienes reciben y quienes administran el tratamiento ignoran cuál es éste, pero no revela lo que ocurre con los evaluadores y los estadísticos que analizan los datos⁴⁴, extremo que debería quedar claro.

A diferencia de lo que ocurre con la asignación oculta (punto 9), el enmascaramiento de los tratamientos no siempre es posible. Aún así, en los ensayos abiertos, hay que justificar las razones por las que no se utilizó. En ocasiones, aunque los tratamientos se hayan enmascarado, sus efectos colaterales son demasiado reveladores. En estos casos, es importante realizar la evaluación a ciegas (lo que suele ser posible, incluso en ensayos abiertos, si el evaluador es una persona distinta de la que trata y controla al paciente) o verificar el éxito del enmascaramiento preguntando a los participantes (individuos que reciben y/o médicos que administran el tratamiento y evalúan el resultado) cuál creen que ha sido el tratamiento administrado^{45,46}. Asimismo, la evaluación a ciegas es especialmente importante cuando las medidas utilizadas son subjetivas.

12. *Métodos estadísticos*. Los métodos estadísticos utilizados para analizar los datos deberían haberse predeterminado durante el diseño del estudio, y es importante indicar si realmente fue así, tanto para el análisis principal como para los secundarios, o bien si se decidieron tras disponer de los datos. La mayoría de los métodos de análisis estadístico proporcionan una estimación puntual del efecto (entendido como una comparación de los tratamientos) y un intervalo de confianza (IC) para la estimación, que usualmente se determina con un nivel del 95%. El IC del 95%, que representa la incertidumbre sobre el efecto estimado, puede interpretarse como el conjunto de valores posibles que son compatibles con los datos del estudio.

Los resultados del análisis pueden expresarse también mediante un grado de significación (p), que expresa la probabilidad de los resultados observados (o más extremos) bajo la hipótesis nula de la correspondiente prueba estadística. Es preferible indicar los valores exactos (p. ej., $p = 0,008$) en lugar de utilizar ciertos puntos de corte (p. ej., $p < 0,05$)⁴⁷.

Las pruebas estadísticas más usuales se basan en la suposición de que las observaciones son independientes. Cuando se dispone de múltiples observaciones para cada individuo (medidas en distintos tiempos o en distintos lugares del cuerpo), las distintas observaciones de un mismo individuo no son independientes, y es necesario utilizar estrategias o métodos de análisis que tengan en cuenta esta dependencia⁴⁸⁻⁵⁰. La dependencia de las observaciones puede ocurrir también en diseños en que los tratamientos se asignan a agregaciones (*clusters*) de individuos (como centros, médicos, colegios o unidades geográficas), pero las observaciones se realizan en cada individuo⁵¹.

Los análisis ajustados deberían incluir solamente variables de ajuste realmente importantes, y señalar si éstas se han preespecificado. En cualquier caso, la decisión de ajustar (o no) no debería basarse en la constatación (o no) de diferencias basales significativas^{52,53}. En cuanto a los análisis de subgrupos, es más recomendable ajustar un modelo lineal que incluya una variable indicador del subgrupo y un término de interacción de ésta con el tratamiento que analizar cada subgrupo por separado⁵⁴.

Resultados

13. *Flujo de participantes*. Para facilitar la comprensión del ensayo¹⁸, el flujo de participantes durante las distintas etapas (inclusión, seguimiento y análisis) debe ilustrarse mediante un diagrama como el de la figura 1. Idealmente debería indicarse el número de individuos inicialmente evaluados para su inclusión en el ensayo (y el de los que no se incluyeron, junto con los motivos) aunque este dato, que permite juzgar la validez externa de los resultados, no siempre está disponible. Es necesario indicar el número total de individuos aleatorizados y, para cada intervención, los números de individuos que se asignaron a la intervención, que la recibieron (tal como estaba previsto), que completaron el estudio según lo estipulado en el protocolo y que se incluyeron en el análisis principal, documentando las sucesivas exclusiones junto con los motivos. Esta información es crítica para juzgar la validez interna de un ECA, ya que las pérdidas y las exclusiones del análisis pueden introducir sesgos en los resultados, especialmente si se dan con diferente frecuencia en los distintos grupos⁵⁵⁻⁵⁷. El diagrama de la figura 1 puede requerir cierta adaptación según cuál sea la naturaleza de la intervención u otras características del ensayo²³.

14. *Reclutamiento*. Las fechas en que se realizó el estudio son importantes para ubicarlo en su contexto histórico. Las fechas de inicio y final del reclutamiento permiten conocer su duración, dato que puede interesar a otros investigadores. En los ensayos con seguimiento de igual duración para todos los pacientes, la fecha de final del seguimiento permite conocer esta duración. En los ensayos en que la variable principal es el tiempo hasta un evento (p. ej., supervivencia), la duración del seguimiento no es la misma para todos los pacientes ya que el ensayo termina en una fecha concreta predeterminada. En estos casos, debe indicarse esta fecha, así como la duración mediana del seguimiento. Si el ensayo se interrumpió como resultado de un análisis intermedio (punto 7), debe indicarse³⁴.

15. Datos basales. Las características demográficas y clínicas basales de los individuos incluidos en un ensayo son importantes para evaluar hasta qué punto los resultados de este ensayo pueden aplicarse a un paciente determinado. Estas características deben describirse para cada uno de los grupos y es recomendable hacerlo en una tabla. Las variables continuas pueden describirse mediante la media y la desviación estándar (mejor que errores estándar o intervalos de confianza) o mediante los tres cuartiles en caso de asimetría marcada. Las variables cualitativas y las cuantitativas discretas con un número limitado de valores pueden describirse mediante frecuencias. A pesar de que es una práctica común, no es útil realizar pruebas formales para comparar los grupos en cuanto a las características basales, y tales pruebas se han desaconsejado^{58,59}. Aunque una asignación aleatoria correctamente implementada asegura la ausencia de sesgos de selección, no garantiza un equilibrio basal perfecto. Si ocurren desequilibrios, serán debidos a fluctuaciones aleatorias (no a sesgo) y las pruebas estadísticas que se utilizarán en el análisis principal sirven, precisamente, para establecer una conclusión teniendo en cuenta las posibles fluctuaciones del azar. Por tanto, asumiendo que la aleatorización es correcta, las decisiones de realizar análisis ajustados por las características basales (p. ej., análisis de covarianza) deben adoptarse en fase de diseño y no a resultados de constatar determinados desequilibrios en ellas^{53,60}.

16. Números analizados. El número de casos incluidos en el análisis es un dato crítico para interpretar su resultado. Este número queda documentado en el diagrama de flujo para el análisis principal, pero puede ser distinto en otros análisis y debe quedar claro cuál fue en todos ellos. Para ello, es muy útil expresar los resultados de las variables dicotómicas en forma de fracción (p. ej., 10/20 en lugar de 50%).

En principio, para evaluar la efectividad de las intervenciones, es aconsejable realizar los análisis por intención de tratar (ITT), es decir, con la inclusión de todos los casos aleatorizados y manteniéndolos en el grupo al que se los asignó, porque así se evita posibles sesgos debidos a pérdidas o exclusiones no aleatorias⁶¹. Sin embargo, la adopción de esta estrategia no siempre es sencilla: si hay pérdidas de pacientes a lo largo del ensayo, la evaluación final prevista no estará disponible, lo que obligará a utilizar técnicas de sustitución de datos ausentes para mantener estrictamente la estrategia de análisis por ITT. Por otra parte, las desviaciones del protocolo en lo que se refiere a la administración de las intervenciones pueden inducir una subestimación del efecto, por lo que es recomendable realizar análisis adicionales que incluyan solamente los casos en que se administró la intervención tal como estaba previsto en el protocolo, lo que se denomina a veces análisis por protocolo (PP)⁶².

A diferencia de lo que ocurre cuando se desea demostrar diferencias entre las intervenciones, cuando el objetivo de un ensayo es demostrar la equivalencia de ambas o la no inferioridad de una respecto de otra, la estrategia por ITT no es conservadora: ciertas desviaciones del protocolo (notablemente las relativas a la administración de la intervención) pueden sesgar los resultados a favor de la hipótesis de equivalencia o no inferioridad⁶³⁻⁶⁵. Por otro lado, la potencia de un análisis PP será inferior a la del análisis por ITT⁶⁶. Por esta razón, en estos ensayos es particularmente importante ofrecer el resultado de ambos análisis, ITT y PP, junto con las definiciones de los conjuntos de individuos que los integran para evitar usos abusivos de esos términos, en particular, el de ITT⁷.

17. Resultados y estimación. Todos los resultados (principal y secundarios) deben describirse para cada grupo de tratamiento (p. ej., medias y desviación estándar o frecuencias), junto con un estadístico adecuado que compare los resultados en ambos grupos (p. ej., diferencia de medias, diferencia de proporciones, riesgo relativo, *odds ratio*, NNT, etc.) y su IC. Aunque los IC se determinan casi siempre para una confianza del 95%, se puede utilizar otros niveles de confianza, pero deben haberse prefijado. Un error frecuente que debe evitarse es ofrecer los IC de los resultados en cada grupo en lugar de hacerlo para el estadístico que los compara. También puede ofrecerse el grado de significación (p) obtenido en una prueba estadística, pero además (no en lugar) del IC^{67,68}.

Es importante facilitar el IC de todos los resultados, principales y secundarios, y no limitarse a los que reflejan diferencias estadísticamente significativas. Con ello se evitará tres fenómenos indeseables: el sesgo de publicación^{69,70}, la errónea pero frecuente tendencia a considerar equivalentes intervenciones que no han mostrado diferencias estadísticamente significativas en ensayos infradimensionados^{71,72}, y la tendencia a evaluar la significación estadística olvidando la relevancia clínica⁷³.

18. Análisis auxiliares. Es inevitable que en un ECA convenga realizar más de un análisis inferencial. Sin embargo, la multiplicidad de estos análisis conlleva un elevado riesgo de resultados falsos positivos, por lo que conviene resistirse a la tentación de analizar demasiadas cosas, especialmente subgrupos de casos^{74,75}, y concentrarse en las cuestiones realmente importantes y previstas en el protocolo o en el plan de análisis. Como ya se ha comentado en el punto anterior, es imprescindible informar acerca de todos los análisis realizados aunque puede ser innecesario ofrecer detalles de todos ellos.

19. Eventos adversos. Además del efecto pretendido, la mayoría de las intervenciones tienen efectos indeseados, y a veces indeseables. El equilibrio entre los beneficios y los riesgos es lo que debe juzgarse para decidir si una intervención es aceptable o útil. Por tanto, es imprescindible facilitar datos sobre los efectos indeseados. Al menos debe describirse la frecuencia de individuos que presentaron acontecimientos adversos (AA) serios (con una definición operativa de lo que se entiende por AA serio⁷⁶), y de los que interrumpieron prematuramente el tratamiento (incluidos los motivos). Cuando un AA se ha presentado en más de una ocasión en el mismo individuo, es útil describir también la frecuencia de los AA.

20. Interpretación. Se ha señalado que, en las publicaciones científicas, la discusión de los resultados suele esmerarse más en justificarlos que en poner de manifiesto posibles limitaciones de estudio^{77,78}. Para evitar esta tendencia, algunas revistas aconsejan estructurar la discusión^{79,80} y presentar: a) un breve resumen de los resultados clave, enfatizando lo que añaden al conocimiento previamente disponible; b) posibles mecanismos y explicaciones de los resultados; c) la comparación con otros trabajos similares, incluido un metaanálisis cuando sea posible; d) las limitaciones del estudio y los métodos adoptados para compensarlas; e) las implicaciones para futuras investigaciones, y f) las implicaciones clínicas.

21. Generalización. La validez externa del estudio, esto es, la posibilidad de generalizar los resultados a conjuntos de individuos más amplios que los incluidos en el ensayo, depende de diversos factores. En primer lugar, de la validez interna, que es un requisito previo para la validez externa: los resultados de un estudio sesgado son inválidos y ya no

procede cuestionarse si son generalizables. Los criterios de selección, que usualmente hacen referencia a la edad, el sexo, la gravedad de la enfermedad, la comorbilidad y las complicaciones, condicionan en gran medida la generalización a otros pacientes. Por ello es imprescindible plantearse seriamente la necesidad y la conveniencia de cada uno de ellos durante el diseño del ensayo, y no imponerlos a menos que exista un buen motivo, ya sea legal, ético o científico. La imitación de estudios previos no es, *per se*, un buen motivo. Además de los criterios de selección, muchos otros aspectos del ensayo determinarán también hasta qué punto pueden generalizarse sus resultados, como el ámbito en que se realizó el ensayo (atención primaria, pacientes hospitalizados, etc.), los métodos utilizados para diagnosticar la indicación del ensayo y para realizar el seguimiento, la intensidad del seguimiento, y la propia intervención o las condiciones en las que se administra. Posiblemente, se evitarían limitaciones en la validez externa de los ensayos si durante su diseño se tuviera muy presente la diferencia entre ensayos explicativos y ensayos pragmáticos^{81,82}, y se decidiera explícitamente el objetivo principal en estos términos. En cualquier caso, decidir si los resultados de un ensayo son aplicables a un paciente concreto es un tarea tan difícil como importante para la práctica clínica⁸³, sobre la que se han desarrollado guías⁸⁴.

22. Evidencia global. Los resultados de un ensayo deben juzgarse en el contexto de la información previa disponible. Aunque lo ideal sería metaanalizar los resultados del ensayo junto con los de ensayos previos similares⁸⁵, esto es poco práctico, muy costoso, y sólo factible si existen tales ensayos. El requisito mínimo es la discusión de los resultados del ensayo en el contexto de la evidencia disponible, sin limitarla a los estudios que obtuvieron resultados similares.

REFERENCIAS BIBLIOGRÁFICAS

- Schulz 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.
- Kjaergård L, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy [Abstract]. In: Abstracts for Workshops and Scientific Sessions, 7th International Cochrane Colloquium, Rome, 1999.
- Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol*. 2000;136:381-5.
- Kjaergård LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology*. 1999;30:1134-8.
- DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med*. 1982;306:1332-7.
- Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272:122-4.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319:670-4.
- Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA*. 1994;272:125-8.
- Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J*. 1977;1:85-7.
- Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317:426-32.
- Altman DG. The scandal of poor medical research [editorial]. *BMJ*. 1994;308:283-4.
- The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA*. 1994;272:1926-31.
- Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. *Ann Intern Med*. 1994;121:894-5.
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637-9.
- Campbell MK, Elbourne DR, Altman DG, for the CONSORT Group. CONSORT statement: extension to cluster randomised trials. *BMJ*. 2004;328:702-8.
- Moher D, Altman DG, Schulz KF, Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *CMAJ*. 2004;171:349-50.
- Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. The CONSORT Group. *JAMA*. 2001;285:1992-5.
- Egger M, Jüni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. The CONSORT Group. *JAMA*. 2001;285:1996-9.
- Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. Consolidated Standards of Reporting Trials. *JAMA*. 1998;279:1487-9.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. *Ann Intern Med*. 2001;134:657-62.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. *JAMA*. 2001;285:1987-91.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. *Lancet*. 2001;357:1191-94.
- Altman DG, Schulz KF, Moher D, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134:663-94.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ*. 1994;309:1286-91.
- World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA*. 1997;277:925-6.
- Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*. 1992;327:248-54.
- Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. National Surgical Adjuvant Breast and Bowel Program. *Pediatric Oncology Group. J Clin Epidemiol*. 1998;51:69-79.
- Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Stat Med*. 1999;18:2605-15.
- McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd ed. New York: Oxford Univ Pr; 1996.
- Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry*. 2000;176:249-52.
- Sanders C, Egger M, Donovan J, Tallon D, Frankel S. Reporting on quality of life in randomised controlled trials: bibliographic study. *BMJ*. 1998;317:1191-4.
- Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994;121:200-6.
- Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics*. 1987;43:213-23.
- Pocock SJ. When to stop a clinical trial. *BMJ*. 1992;305:235-40.
- Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med*. 1985;4:521-6.
- Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does [editorial]. *BMJ*. 1998;317:362-3.
- Martyn C. Not quite as random as I pretended. *BMJ*. 1996;347:70.
- Williams DH, Davis CE. Reporting of assignment methods in clinical trials. *Control Clin Trials*. 1994;15:294-8.
- Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274:1456-8.
- Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med*. 1987;6:315-28.
- Haag U. Technologies for automating randomized treatment assignment in clinical trials. *Drug Information Journal*. 1998;32:11.
- Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetsis E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*. 1994;44:16-20.
- Götzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials*. 1996;17:285-90; discussion 290-3.
- Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285:2000-3.
- Prasad AS, Fitzgerald JT, Bao B, Beck FW, Chandrasekar PH. Duration of symptoms and plasma cytokine levels in patients with the common cold treated with zinc acetate. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2000;133:245-52.
- Quitkin FM, Rabkin JG, Gerald J, Davis JM, Klein DF. Validity of clinical trials of antidepressants. *Am J Psychiatry*. 2000;157:327-37.
- Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. En: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2nd ed. London: BMJ Books; 2000. p. 171-90.
- Altman DG, Bland JM. *Statistics notes. Units of analysis*. *BMJ*. 1997;314:1874.

49. Bolton S. Independence and statistical inference in clinical trial designs: a tutorial review. *J Clin Pharmacol*. 1998;38:408-12.
50. Greenland S. Principles of multilevel modelling. *Int J Epidemiol*. 2000; 29:158-67.
51. Kerry SM, Bland JM. Analysis of a trial randomised by clusters. *BMJ*. 1998;316:54.
52. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355:1064-9.
53. Altman DG. Adjustment for covariate imbalance. En: Armitage P, Colton T, editors. *Encyclopedia of Biostatistics*. Chichester: John Wiley; 1998. p. 1000-5.
54. Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ*. 1996;313:808.
55. Altman DG. Randomisation [editorial]. *BMJ*. 1991;302:1481-2.
56. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med*. 1979;301:1410-2.
57. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation*. 1981;64:669-73.
58. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet*. 1990;335:149-53.
59. Senn S. Base logic: tests of baseline balance in randomized clinical trials. *Clinical Research and Regulatory Affairs*. 1995;12:171-82.
60. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med*. 1991;10:1595-605.
61. Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000;21:526.
62. Nagelkerke N, Fidler V, Bensen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med*. 2000;19:1849-64.
63. ICH E9. Statistical principles for clinical trials, february 1998 [accedido el 20 de Julio 2005]. Disponible en: <http://www.ich.org>
64. Chadwick D. Monotherapy comparative trials: equivalence and differences in clinical trials. *Epilepsy Res*. 2001;45:101-3.
65. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ*. 1996;313:36-9.
66. Brittain E, Lin D. A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Stat Med*. 2005;24(1):1-10.
67. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*. 1986; 292:746-50.
68. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med*. 1997;126:36-47.
69. Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics*. 2000;49:359-70.
70. Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med*. 2000;19:3325-36.
71. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311:485.
72. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 «negative» trials. *N Engl J Med*. 1978;299:690-4.
73. Luus HG, Muller FO, Meyer BH. Statistical significance versus clinical relevance. Part II. The use and interpretation of confidence intervals. *S Afr Med J*. 1989;76(11):626-9.
74. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355: 1064-9.
75. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116:78-84.
76. Ioannidis JP, Lan J. Completeness of safety reporting in randomized trials. An evaluation of 7 medical areas. *JAMA*. 2001;285:437-43.
77. Horton R. The rhetoric of research. *BMJ*. 1995;310:985-7.
78. Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA*. 1998; 280:227-8.
79. Docherty M, Smith R. The case for structuring the discussion of scientific papers [editorial]. *BMJ*. 1999;318:1224-5.
80. Information for authors. *Annals of Internal Medicine* [accedido el 22 Julio 2005]. Disponible en: <http://www.annals.org>
81. Schwartz D, Flamant R, Lellouch J. *L'essai thérapeutique chez l'homme*. Flammarion Médecine-Sciences; 1970.
82. Roland M, Torgerson DJ. Understanding controlled trials: what are pragmatic trials? *BMJ*. 1998;316:285.
83. Davey Smith G, Egger M. Who benefits from medical interventions? [editorial]. *BMJ*. 1994;308:72-4.
84. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA*. 1998;279:545-9.
85. Gøtzsche PC, Gjørup I, Bonne'n H, Brahe NE, Becker U, Burcharth F. Somatostatin v placebo in bleeding oesophageal varices: randomised trial and meta-analysis. *BMJ*. 1995;310:1495-8.