

Cómo mejorar la calidad de los ensayos clínicos

Inés Galende Domínguez^a, José A. Sacristán^b y Javier Soto^c

^aDirección General de Farmacia y Productos Sanitarios. Ministerio de Sanidad y Consumo. Madrid.

^bDepartamento de Investigación Clínica, Lilly S.A.

^cUnidad de Farmacología Clínica. Hospital Santa Cruz. Liencres. Cantabria.

Es evidente que el paso previo necesario para mejorar algo es conocer sus deficiencias y a ser posible el alcance de las mismas, puesto que si éstas no tienen una repercusión importante, quizá no merezca la pena el esfuerzo de intentar corregirlas.

En el caso concreto de la realización de ensayos clínicos (EC), hemos detectado tanto la falta de descripción de determinados aspectos importantes a la hora de validar los resultados o admitir las conclusiones que se extraen, como errores conceptuales que hacen dudar de la veracidad del estudio.

Para algunos autores uno de cada 2 EC dado a conocer de forma resumida nunca será publicado con el detalle suficiente para que pueda valorarse adecuadamente su validez¹, hecho que es agravado por la no publicación de los EC con resultados negativos².

La responsabilidad de estos hechos debe recaer no sólo sobre los investigadores, sino también sobre los comités de ética de la investigación, patrocinadores y los directores de revistas científicas. Todos deben intentar reducir tanto la publicación incompleta como la no publicación de los EC bien diseñados independientemente de que los resultados obtenidos sean o no los esperados³.

Dado que el empleo de un determinado fármaco en la práctica asistencial muchas veces se basa en los EC publicados en las revistas científicas, es preciso que éstos estén diseñados y realizados y que su publicación sea completa y detallada para que no dé lugar a errores de interpretación. Por tanto es de gran importancia tanto para el médico asistencial como para sus pacientes (y principalmente para éstos), que no se produzcan errores de prescripción por una incorrecta interpretación de los resultados de estudios publicados, ya que sus consecuencias podrían ser

no sólo la administración de tratamientos ineficaces o peligrosos, sino también demorar innecesariamente otros tratamientos que podrían ser eficaces para determinados pacientes.

A continuación indicaremos cuáles son los errores más frecuentemente encontrados en la literatura por nuestro grupo⁴ en el diseño, conducción y publicación de los EC al aplicar una lista guía propuesta previamente⁵. Ello permite conocer de forma más o menos exacta los errores cometidos con mayor frecuencia en España, en el diseño y realización de los EC. Finalmente, analizaremos cada uno de ellos intentando ofrecer soluciones a los problemas planteados.

Comparabilidad de los grupos de tratamiento

En nuestro estudio hemos detectado que en el 70% de los casos, como término medio entre las 3 revistas seleccionadas, no se especifica si los grupos son comparables respecto a sus características basales y factores pronósticos conocidos. Tampoco se describe detalladamente si la asignación a los grupos de tratamiento se hizo de forma aleatoria, ni en caso afirmativo, el método seguido. Estas deficiencias también han sido señaladas por otros autores^{6,7}. Este hecho es de gran importancia puesto que la asignación aleatoria de los sujetos a los diferentes grupos de tratamiento evita los posibles *sesgos de selección*. Así, la única causa que pudiera justificar las diferencias encontradas, si existe, sería el tratamiento recibido. Por tanto, protegería frente a posibles factores, conocidos o desconocidos, y otros sesgos accidentales a lo largo del estudio, que pudieran interferir en la respuesta⁸.

La asignación aleatoria se define como el procedimiento sistemático y reproducible por el que los sujetos participantes en un EC son distribuidos al azar en los distintos grupos de tratamiento⁹.

Las características que debe cumplir un buen método de aleatorización¹⁰ se especifican en la

tabla I. Ello requiere, en primer lugar, que la asignación aleatoria se realice de forma "ciega", es decir, desconocida para el investigador.

Un error detectado con frecuencia es describir como asignación aleatoria la inclusión de pacientes en uno u otro grupo en función de su orden de reclutamiento. En este caso sería el azar el que determinaría que un sujeto fuera antes que otro a la consulta, pero podría estar sesgado, por ejemplo, porque determinados días se cite a pacientes con características especiales como nuevos diagnósticos u otros. Este modo de aleatorización es incorrecto por varias razones: en primer lugar no es impredecible, pero tampoco cumple otras propiedades como ser reproducible y estar basado en un modelo matemático¹¹. Por las razones descritas tampoco pueden considerarse adecuados los métodos de lanzar una moneda al aire o días pares o impares.

Tampoco es correcto que el investigador encargado de seleccionar a los pacientes disponga de una tabla donde conste la asignación de todos los sujetos a los grupos de tratamiento, pues al conocer a qué grupo va a ser asignado el siguiente sujeto, puede producirse un sesgo de selección (intentar incluir a los sujetos con determinadas características, p. ej., patología más grave, en uno de los grupos, tanto de forma consciente como inconsciente).

La forma más frecuente de realizar la aleatorización ciega consiste en elaborar un sobre, opaco para que no pueda verse al trasluz, en el que se incluye el tratamiento al que ha sido asignado el paciente. Este sobre cerrado, que en su parte externa indica el número del sujeto según el orden de inclusión en el estudio, se entrega al investigador y es abierto una vez que el sujeto ha dado su conformidad para participar en el ensayo. Si el estudio fuese doble ciego, no podría abrirse hasta que el ensayo hubiera finalizado.

Métodos de distribución aleatoria

Aunque para cada diseño o protocolo de ensayo clínico será más adecuado un determinado método de distribución aleatoria, y sin querer entrar en una descripción detallada de cada uno de ellos, se describen a continuación sus principales ventajas e inconvenientes:

Aleatorización simple. Es el método más sencillo de asignación aleatoria. Supongamos un EC en el que vamos a estudiar 2 tratamientos distintos A y B. Si en una tabla de números aleatorios (fig. 1) asignamos el tratamiento A a los números 1-4 y el tratamiento B a los números 5-9, co-

TABLA I
CARACTERÍSTICAS DE UN BUEN
ESQUEMA DE ALEATORIZACIÓN

La asignación a los grupos de tratamiento debe ser desconocida e impredecible
El proceso de asignación debe ser reproducible y estar documentado
Debe basarse en propiedades matemáticas conocidas
No debe permitir que se prevenga la secuencia de los tratamientos
Los fallos en la asignación de tratamientos pueden ser detectados

menzando por un punto determinado, que puede ser un extremo o cualquier otro de la tabla y siguiendo un orden prefijado (vertical, horizontal), sólo tenemos que sustituir el número por el tratamiento que corresponda (A o B).

Los inconvenientes que presenta este método¹¹ son los siguientes:

1. El distinto tamaño de los grupos de tratamiento, sobre todo si el número total de sujetos es inferior a 100 (cuanto menor es el número de sujetos mayor es la probabilidad de desigualdad), lo que a su vez incide en la significación estadística.

2. En los ensayos clínicos multicéntricos (ECM), la distribución de los grupos de tratamiento en cada centro puede no estar equilibrada.

3. No se hacen análisis intermedios, por la misma razón.

4. En caso de existir factores pronósticos, éstos pueden no estar distribuidos de forma homogénea y, por tanto, dificultar la interpretación de los resultados.

Asignación aleatoria balanceada o restrictiva. Asegura igual número de sujetos en cada grupo de tratamiento, con lo que facilita el análisis de los resultados. A su vez puede realizarse por distintos métodos:

1. **Aleatorización balanceada global.** Se realiza igual que la asignación aleatoria simple pero en ella se asegura que el número de sujetos de cada grupo de tratamiento, al final del estudio, es igual. Presenta los mismos inconvenientes que la simple en cuanto al análisis intermedio y en los ECM.

2. **Aleatorización por bloques.** Se puede realizar según el tamaño de los bloques: a) asignando la secuencia AB a los números 0-4, y BA a los

11	19	15	5	9	0	6	13	7	2	16	1	12	18	4	17	10	8	3	14
14	12	0	1	19	8	7	17	11	18	2	15	5	9	4	16	10	6	13	3
5	17	2	4	16	19	10	11	14	7	12	15	1	18	6	9	0	3	13	8
8	13	3	12	10	5	17	2	6	7	16	19	0	1	4	11	14	15	18	9
11	6	8	0	1	10	13	18	12	14	17	7	4	5	3	9	19	16	2	15
17	18	3	6	9	15	14	5	4	19	2	1	0	8	11	13	7	12	10	16
14	19	13	16	1	9	18	0	5	15	4	12	10	11	2	3	8	6	7	17
0	3	2	13	7	8	19	12	5	9	16	6	4	17	15	14	1	11	18	10
11	19	2	6	12	15	17	0	10	3	4	14	7	5	16	13	1	8	9	18
13	18	9	6	5	17	19	0	8	10	15	7	11	3	12	4	16	1	2	14
9	3	4	17	18	2	13	14	15	11	0	8	1	7	6	19	16	12	10	5
2	9	17	12	6	19	14	4	1	7	5	3	10	13	0	18	8	15	11	16
14	8	11	2	13	6	5	0	10	12	19	4	16	15	9	17	7	18	1	3
5	9	11	3	7	14	19	15	0	17	2	12	18	4	13	16	10	1	6	8
3	19	11	17	18	10	6	4	14	2	1	16	9	5	7	8	12	13	0	15
13	3	9	11	2	7	0	15	19	4	14	10	12	6	5	1	17	16	18	8
0	5	18	12	3	11	8	15	6	16	9	4	7	2	19	17	14	10	1	13
10	15	0	16	7	5	4	13	12	1	17	3	9	14	11	8	6	18	2	19
2	16	13	19	8	6	17	9	14	4	12	3	1	11	5	15	0	10	7	18
18	15	5	11	6	3	14	13	7	0	9	17	2	1	8	10	12	19	4	16
9	14	3	6	16	1	0	11	4	2	10	12	19	13	7	15	18	8	17	5
15	18	4	12	1	7	11	10	5	17	14	8	2	0	3	6	9	19	13	16
12	1	7	13	19	8	6	4	10	14	0	18	15	9	17	16	11	2	5	3
15	11	3	10	14	9	16	2	5	17	18	19	4	6	13	1	8	0	7	12
1	17	16	10	15	18	0	7	11	9	2	14	3	5	13	12	6	4	19	8
5	9	16	12	6	17	19	15	2	14	11	0	3	10	8	18	1	4	7	13
9	8	0	7	4	17	19	3	5	6	13	15	16	10	11	12	1	14	2	18
9	2	17	7	16	14	5	15	19	8	13	6	0	4	18	3	10	11	1	12
9	4	14	1	5	0	6	10	15	17	8	16	19	18	7	2	11	13	3	12
6	4	17	14	16	2	1	8	15	11	3	0	10	18	5	13	19	7	12	9
11	6	14	13	10	4	7	18	19	12	15	2	8	5	17	3	1	16	9	0
17	2	14	8	4	11	9	12	3	18	6	13	1	19	7	0	16	5	10	15
1	11	5	9	4	17	14	7	6	12	0	10	19	15	8	16	3	13	2	18
8	19	5	15	9	14	4	1	18	16	11	0	3	12	17	13	7	10	2	6
12	11	6	18	7	13	3	2	14	19	10	9	16	0	4	15	5	8	17	1
17	7	11	4	3	15	16	9	8	0	5	18	10	19	2	13	12	14	1	6
7	11	18	0	17	19	15	12	10	5	8	3	9	13	4	14	1	2	16	6
5	8	0	2	3	13	15	19	6	18	1	10	9	12	14	16	4	7	17	11
15	11	1	12	7	14	13	19	2	16	10	6	18	8	3	17	0	5	4	9
8	1	9	10	6	15	4	19	0	18	2	7	16	13	5	3	11	14	17	12
13	1	17	14	11	16	3	5	7	9	0	15	19	6	18	12	4	10	8	2
3	12	9	4	6	15	5	16	17	18	7	2	19	11	14	8	1	13	10	0
10	3	8	15	2	16	19	4	1	5	13	14	6	7	11	0	17	12	18	9
16	5	9	1	15	18	17	12	10	19	8	13	6	11	4	14	7	3	0	2
19	1	9	16	3	11	8	15	4	13	12	18	0	10	7	6	2	14	5	17
0	10	3	5	13	17	19	8	7	16	14	9	11	12	4	6	2	18	15	1
3	16	15	13	7	9	0	2	18	14	5	10	17	4	19	11	12	1	8	6
14	16	15	7	4	17	2	10	3	1	8	11	18	0	19	12	6	13	5	9
3	1	17	18	19	0	5	9	14	10	8	2	15	4	12	6	16	7	13	11
11	17	13	19	16	18	2	15	1	8	7	10	14	6	0	9	4	5	3	12

Fig. 1. Tabla de números aleatorios.

números 5-9 de una tabla como la de la figura 1, o b) generando tablas del tamaño de los bloques y asignando a cada número un tratamiento. Por ejemplo, si el bloque es de 10 y los grupos de tratamiento son 2 (A y B), se generan números aleatorios del 0 al 9 tantas veces como bloques que-

ramos y se asignan del 0-4 al grupo A, y del 5-9 al grupo B.

Incluir en cada bloque el mismo número de sujetos asignados a cada grupo de tratamiento permite no sólo el análisis final, sino análisis parciales y, por ejemplo, por hospital, en los ECM.

Es particularmente importante en los casos en los que las características de los pacientes pueden variar a lo largo del tiempo como, por ejemplo, cambios en la política de reclutamiento¹².

La longitud de los bloques debe ser ni muy corta (lo que permitiría adivinar la secuencia de tratamientos), ni muy larga (que pueda desequilibrar los grupos en análisis intermedios), y siempre múltiplo del número de tratamientos en estudio.

Generalmente se acepta que los bloques sean de 8-10 y en cualquier caso que el investigador desconozca la longitud de los mismos^{10,13}.

3. Aleatorización sistemática. La asignación se realiza según la secuencia generada para un primer bloque, que se repite tantas veces como fuera necesario según el número de sujetos a incluir en el estudio. No es recomendable por los sesgos de selección a que da lugar.

Estratificación. Consiste en la asignación de los sujetos a los grupos de tratamiento, pero según unos factores pronósticos o determinadas características especiales que pueden influir en los resultados del estudio.

Las ventajas que aporta son, por un lado, evitar la atribución de las diferencias observadas al tratamiento, cuando los resultados pueden deberse única y exclusivamente a estas variables o factores pronósticos no considerados y, por otro lado, la estratificación *a priori* aumenta el poder estadístico del estudio, sobre todo si el número de sujetos es inferior a 50¹³.

No obstante, es preciso tener en cuenta que si estratificamos por un gran número de variables aumenta ostensiblemente la complejidad, y la eficacia disminuye, al encontrarnos con un número insuficiente de sujetos en cada estrato.

Minimización. La minimización es una forma especial de estratificación en la que se consideran todas las variables que se conoce que pueden influir en los resultados del estudio. Para ello se calcula la medida de desequilibrio sobre los factores pronósticos del nuevo paciente, y se le asigna al grupo que minimiza el desequilibrio global¹¹.

Siempre precisa de una aleatorización centralizada, y requiere una entrada lenta de los sujetos en el estudio para poder ir equilibrando los grupos de tratamiento, a partir de los factores pronósticos conocidos.

Consentimiento aleatorizado. El modelo propuesto por Zelen¹⁴ de asignar aleatoriamente a los pacientes antes de haber solicitado su consentimiento para participar en un ensayo, no ha tenido

aceptación debido a que, aunque se aducía que así no se alteraba la relación médico-paciente, no consigue el objetivo principal de la aleatorización que es evitar los sesgos de selección, además de dificultar enormemente la interpretación de los resultados. Tampoco se considera satisfactorio desde el punto de vista ético, puesto que a los sujetos que les corresponde aleatoriamente el tratamiento estándar (grupo control) no se les solicita el consentimiento ni se les informa que participan en un EC. Este modelo también impide la realización de los EC con carácter doble ciego.

Otros tipos de asignación aleatoria

1. Centralizada: se realiza por teléfono o fax, y consiste en asignar a los pacientes a cada grupo de tratamiento según el orden de inclusión, independientemente del centro o país donde se esté realizando el EC.

2. Proporcional: este tipo de asignación aleatoria no balanceada incluye generalmente el doble de sujetos en el grupo experimental que en el control. Aunque pueden utilizarse otras proporciones (3:1, 4:1) debe tenerse en cuenta que a medida que son mayores el poder estadístico del estudio disminuye, y en la práctica suelen ser inaceptables¹⁰. Su fin es obtener una mayor experiencia con el fármaco en estudio, sobre todo en las primeras fases de la investigación.

Medición objetiva de los resultados

Casi el 80 % de los estudios publicados analizados carecían de información acerca del enmascaramiento de los tratamientos. Así como la aleatorización es imprescindible en un EC controlado, el enmascaramiento puede ser obligado, deseable o inaceptable desde el punto de vista ético.

Las circunstancias que hacen obligado un enmascaramiento de los tratamientos son aquellas que precisan de una valoración subjetiva como, por ejemplo, el dolor. En general, los ensayos clínicos cuya variable principal de respuesta (eficacia o seguridad) sea subjetiva y, por tanto, el conocimiento del tratamiento recibido pueda interferir de forma importante con su valoración, deben ser enmascarados. Otros ensayos en los que la variable pueda ser objetiva u objetivable (un parámetro bioquímico o microbiológico), en los que el enmascaramiento del tratamiento sea extremadamente difícil desde el punto de vista práctico (fármacos de administración parenteral), o que dé lugar a serios problemas éticos (fármacos muy tóxicos), pueden ser susceptibles de otras

medidas como la del observador ciego, es decir, la persona que evalúa la respuesta desconoce la asignación a los grupos de tratamiento. Además, debe tenerse en cuenta que aunque 2 fármacos pueden formularse para ser indistinguibles en forma de cápsulas, las modificaciones de la forma galénica y quizá de las características farmacocinéticas y/o farmacodinámicas pueden hacer necesarios estudios de bioequivalencia.

La descripción y la valoración de la eficacia del enmascaramiento (mediante preguntas al sujeto como ¿conoce Vd. el fármaco que está tomando?, ¿por qué?) deben ser incluidas en toda publicación, puesto que los estudios no enmascarados suelen sobrevalorar la eficacia del fármaco experimental y, en especial, cuando el control utilizado es un placebo. El enmascaramiento o carácter ciego de un EC es toda una serie de medidas o precauciones que se toman con el fin de que, a lo largo del estudio, bien el paciente, el médico, o ambos, desconozcan la asignación de los tratamientos. Incluye asignación aleatoria ciega, similares características de las muestras e iguales etiquetas.

El objetivo perseguido al enmascarar los tratamientos utilizados en un EC es evitar los *sesgos de valoración* por parte de todos los implicados en el ensayo, pero principalmente el paciente y el investigador.

1. El paciente. El hecho de participar en un EC en el que aleatoriamente se le ha asignado un tratamiento puede dar lugar, por un lado a un efecto positivo (*efecto placebo*) en aquellos pacientes a los que le ha correspondido el tratamiento que consideran más eficaz, o un efecto negativo (*efecto nocivo*) en los pacientes a quienes les ha correspondido el tratamiento no deseado. Ambos efectos impiden una valoración objetiva del efecto del tratamiento^{10,13}.

También es frecuente que si el paciente conoce el tratamiento que le corresponde, y no es el deseado, se retire del estudio para ser tratado con las medidas habituales (*sesgo de retirada*).

2. El investigador. Si las variables de valoración de la eficacia no son totalmente objetivas, el investigador que conoce la asignación de tratamientos puede tener una tendencia tanto consciente como inconsciente a favorecer uno de los tratamientos en estudio (*sesgo de evaluación*). Esto puede ocurrir con variables "subjetivables", como la tensión arterial o la interpretación de imágenes radiológicas. El momento de aparición de este sesgo puede ser tanto al inicio del estudio como a lo largo del mismo si existe una diferencia que se manifiesta clínicamente¹¹.

También el investigador puede incurrir en *sesgo de retirada* si "a su juicio" actúa en el mejor interés del paciente retirándolo del ensayo. Ello ocurre con mayor frecuencia cuando uno de los grupos del ensayo está recibiendo placebo.

Para conseguir que un ensayo esté bien enmascarado y, por tanto, para evitar los sesgos anteriormente descritos, deben considerarse las siguientes recomendaciones¹⁰:

1. La asignación aleatoria debe cumplir los criterios anteriormente expresados.

2. Los tratamientos deben estar enmascarados en el mayor grado posible.

3. La recogida de datos o la evaluación de la respuesta deben ser realizadas por personas que desconozcan la asignación de tratamientos.

4. Si la variable principal de valoración de la respuesta es subjetiva (dolor), o "subjetivable" (TA), debe considerarse la posibilidad de un observador imparcial, tercer observador o evacuador ciego.

5. Nunca deben enmascararse los tratamientos si por ello sometemos a algún tipo de riesgo a los sujetos.

Tipos de ciego

Simple ciego. Generalmente se refiere a que el sujeto desconoce el grupo de tratamiento al que pertenece, aunque es a veces el investigador quien desconoce la asignación de tratamientos⁹.

Es imprescindible que en los EC en los que la variable principal de valoración es evaluada por el paciente, por ejemplo, la intensidad del dolor en una patología dada, al menos este último desconozca el tratamiento que se le está administrando.

Doble ciego. Este tipo de enmascaramiento indica que ni el sujeto del ensayo ni el investigador conocen el tratamiento administrado⁹. De gran importancia cuando la variable principal de valoración por parte del médico y/o paciente es subjetiva (mejoría de los síntomas) o "subjetivable".

Triple ciego/cuádruple ciego. En estos EC, además del sujeto y del investigador, también el monitor desconoce la asignación de los tratamientos. Se denomina cuádruple ciego cuando el análisis estadístico se realiza sin desvelar los códigos de los tratamientos estudiados. Estos términos no son aceptados de forma universal.

Doble simulación. Es un procedimiento para mantener el carácter doble ciego, que consiste

en la administración de un placebo de características externas iguales a las de uno o varios de los tratamientos activos, en situaciones en las cuales la forma de presentación y/o pauta posológica (dosis, vía de administración y frecuencia) sea diferente a la del grupo control⁹.

Este término se corresponde con el inglés de *double-dummy*, y a veces ha sido traducido como doble engaño. Existen estudios en los que la complejidad de realizar una doble simulación ocasiona tantos problemas técnicos y/o prácticos que justificaría que el estudio no fuera enmascarado a pesar de que las variables no fueran totalmente objetivas. Un ejemplo podría ser la administración de 2 antibióticos, uno cada 6 h y otro cada 8 h. Para aplicar este procedimiento exigiría la toma de 2 formulaciones distintas cinco veces al día, hecho bastante complejo que además repercutiría negativamente en el cumplimiento terapéutico.

Pérdidas

También en aproximadamente un 80 % de los artículos revisados no aparecen datos referentes a las pérdidas de los sujetos ni antes ni después de la aleatorización.

La importancia que este tipo de pérdidas tiene en relación con los resultados del ensayo depende de cuándo éstas se produzcan y las medidas tomadas para corregirlas.

Las pérdidas de los sujetos elegibles producidas antes de su inclusión en el ensayo (tanto en la fase de reclutamiento como durante los períodos de lavado o de estabilización) condicionan fundamentalmente la capacidad de generalización de los resultados. En determinadas circunstancias sólo un subgrupo de los sujetos que cumplen los criterios de selección pueden elegir participar en el estudio, por ejemplo, los que presentan una patología más grave¹⁵. Por ello deben recogerse, al menos, las características basales de estos sujetos y posteriormente compararlas con las de los sujetos aleatorizados. Como norma se considera que si un estudio incluye entre el 80 y el 90 % de los sujetos elegibles es representativo de la población estudiada.

Las pérdidas posteriores a la aleatorización, sin embargo, tienen un significado diferente. Éstas pueden ser derivadas de la aparición de acontecimientos adversos, retirada del consentimiento, falta de eficacia, etc. En este sentido, la práctica habitual de considerar la última observación clínica antes de la retirada del sujeto del estudio se considera incorrecta puesto que ésta puede haberse producido por múltiples causas. Estos ca-

sos deben ser descritos individualizadamente en el informe final, presentando además un análisis descriptivo de todas las posibles causas, momento en que se produjeron, dosis de fármaco tomado hasta el momento, de la retirada y/o abandono.

En principio, las retiradas o abandonos no deben suponer más del 20 % de los sujetos puesto que podrían alterar la certeza de los resultados, y en todo caso deben estar previstas cuando se calcula el número de pacientes necesarios para demostrar la hipótesis de trabajo.

Un error bastante frecuente consiste en sustituir a los sujetos que han sido retirados o han abandonado el ensayo, asignándoles el mismo tratamiento que les había correspondido a éstos. Lo correcto es reemplazarlos, elaborando una lista de aleatorización que contemple estas posibles pérdidas y a medida que se retira un paciente asignarle el siguiente número de la lista de modo ciego, y así hasta conseguir el total necesario para el análisis de sujetos evaluables. En cualquier caso es aconsejable realizar el seguimiento de todos los sujetos hasta que finalice el ensayo¹⁶.

Diseño estadístico

La deficiencia más notoria en todos los artículos revisados (más del 95 %), es la descripción del diseño estadístico del ensayo. No se especifica por qué se ha elegido el número determinado de sujetos, ni las razones para elegir un tipo de diseño (en paralelo o cruzado) o aleatorización (en bloques o estratificación) o las diferencias estadísticas que se consideran relevantes desde el punto de vista clínico.

La eficacia de un nuevo fármaco puede ser demostrada comparándolo bien con un placebo o bien con otro tratamiento activo. La superioridad en ambos casos debe establecerse a partir de la demostración de una diferencia clínicamente aceptable o relevante. Esta diferencia debe ser considerada para formular la hipótesis alternativa.

En otros casos, y debido a que existe un tratamiento estándar de eficacia probada para la patología en estudio, bastaría con que la comparación demostrara equivalencia terapéutica, en vez de superioridad. La clásica hipótesis nula no sería aceptable en esta situación¹².

La importancia que tiene predeterminar el tamaño de la muestra, en función de una diferencia clínicamente significativa, es tal, que equivale a creernos o no creernos los resultados. Es decir, si no se incluye un suficiente número de

sujetos en el ensayo no puede asegurarse que no existan diferencias entre los grupos, ya que lo más probable es que el estudio no tenga el poder suficiente para demostrarlas, “la ausencia de prueba no es una prueba de ausencia”.

Para calcular *a priori* el número de sujetos que se necesitan para demostrar con una probabilidad determinada que existen o no diferencias estadísticamente significativas entre los grupos de tratamiento se deben considerar una serie de factores: la probabilidad de un resultado falso positivo o error alfa (arbitrariamente inferior al 5%), la probabilidad de un resultado falso negativo o error beta (generalmente inferior a un 20%), la diferencia clínicamente significativa que se espera encontrar y las posibles pérdidas después de la aleatorización a lo largo del estudio.

Puesto que existen muchas fórmulas para calcular el tamaño de la muestra, debe especificarse claramente cuál se ha utilizado, así como las variables que se han considerado para su cálculo (media, tanto por ciento de respuestas, diferencia relevante desde el punto de vista clínico), que deben estar basadas en los resultados de estudios previamente realizados o publicados.

Desde un punto de vista práctico, en la publicación de los resultados de un ensayo clínico, podría corregirse parcialmente esta deficiencia si se calculara el poder del estudio *a posteriori*, así se podría valorar si los resultados obtenidos contestan, y en qué medida, a la hipótesis inicial.

Otros puntos que deben considerarse en el diseño de un ensayo incluyen el tipo de distribución estadística que siguen las principales variables de respuesta. Así, si la variable ofrece unos resultados que producen una distribución de carácter discreto, se recomienda utilizar la distribución exacta.

Otros factores que también deben considerarse de forma explícita dependen del propio ensayo: en los ensayos multicéntricos debe tenerse en cuenta la posible influencia de la variabilidad intercentros en la tasa de reclutamiento, número y características de los sujetos asignados a los posibles grupos de tratamiento, etc.; en los ensayos de diseño cruzado, frecuentemente utilizados en los estudios de bioequivalencia en voluntarios sanos, deben valorarse los posibles fenómenos de solapamiento (*carry over*) y efecto período.

Análisis de los resultados

En más del 80% de los ensayos revisados no se incluían datos suficientes sobre el análisis estadístico de los resultados, no se especificaba si se habían realizado análisis por subgrupos o se había valorado la repercusión de las pérdidas.

Comentario aparte merece el concepto de “análisis por intención de tratar”. Por definición, este análisis incluye a todos los sujetos que han sido aleatorizados a los grupos de tratamiento, independientemente de que hayan recibido o no alguna dosis del tratamiento asignado o se haya producido algún error en la administración, es por *intención de tratamiento*, no por tratamiento administrado. Modificaciones de este concepto se refieren también a todos aquellos pacientes de los que se ha obtenido al menos una evaluación de variable principal de eficacia después de la medición basal, o a todos los pacientes que han sido distribuidos de forma adecuada y cumplen todos los criterios de inclusión y ninguno de exclusión, o finalmente a todos los pacientes correctamente distribuidos que padecen la enfermedad en estudio y han recibido al menos una dosis del tratamiento¹². Su importancia reside en que nos da una idea de lo que sucederá en la práctica clínica habitual (efectividad) por comparación con lo que ocurre en las condiciones de un ensayo, que es lo que determina el análisis “por tratamiento”. Este último, también denominado “por protocolo”, “de casos válidos” o “de eficacia” incluye sólo a los pacientes evaluables, es decir, se ha conseguido medir la variable principal de eficacia en las circunstancias y tiempos previstos, se ha tomado una dosis determinada o han completado como mínimo unos días de tratamiento y no ha habido ninguna violación del protocolo. Siempre que existan discrepancias en los resultados de los análisis anteriores debe darse una explicación para las mismas, y en cualquier caso, llevar a cabo la descripción detallada de los sujetos que se excluyeron del análisis por intención de tratar.

El análisis de seguridad debe incluir a todos los sujetos que han tomado al menos una dosis de tratamiento.

Otro error frecuentemente detectado es la realización de análisis por subgrupos *a posteriori*, no planificados y, por tanto, no ajustados al diseño del estudio. Los resultados que se obtienen de este modo no sólo pueden ser incorrectos, sino que su generalización puede dar lugar a graves errores.

Es preciso, por tanto, indicar cuáles son las pruebas estadísticas utilizadas para el análisis de los resultados, si éstas son las previstas inicialmente, y si se han obtenido los valores que permitan concluir diferencias estadísticamente significativas para una o 2 colas o mediante intervalos de confianza. En determinados casos, análisis intermedios o diseño secuencial, será preciso, además, ajustar el error tipo alfa.

Recogida de acontecimientos adversos

Más del 90 % de los artículos no contenían información sobre este aspecto. Aunque ello no condiciona la credibilidad de los resultados y/o la generalización de los mismos, es importante puesto que su conocimiento permite valorar la relación riesgo/beneficio del/de los fármacos en estudio y su posición frente a otras alternativas disponibles.

A pesar de que los ensayos clínicos rara vez incluyen un número de sujetos suficiente para poder detectar reacciones adversas graves infrecuentes, la conclusión de que un fármaco no presenta problemas de seguridad no puede afirmarse por el hecho de que no se haya podido demostrar diferencias en los perfiles de seguridad entre los fármacos comparados.

El análisis de seguridad debe realizarse en todos los pacientes que han tomado al menos una dosis del fármaco y se dispone de información sobre los resultados.

Es deseable que se enumeren los acontecimientos adversos, entendiendo como tales cualquier experiencia no deseada que le ocurra a un sujeto, se considere o no relacionada con él o los productos en investigación, haciendo hincapié en los más graves y más frecuentes. Debería indicarse también el método utilizado para su detección (notificación espontánea o pregunta orientada) y el algoritmo seguido para establecer su relación de causalidad.

Aspectos éticos

Tanto la aprobación del ensayo por un comité específico –Comité de Ensayos Clínicos– como la obtención del consentimiento informado de los sujetos participantes son requisitos legales exigidos en España desde 1978¹⁷ y, sin embargo, en el 90 y el 70 % de los casos, respectivamente, no se hacía mención en los ensayos publicados incluidos en nuestra revisión. También las normas de Buena Práctica Clínica consideran estos aspectos, junto con otros aplicables a la investigación biomédica, que será preciso observar para el reconocimiento de los resultados de la investigación. El cumplimiento de estas normas se exige en España desde agosto del año pasado al entrar en vigor, 3 meses después de su publicación en el BOE, el Real Decreto 561/1993, de 16 de abril, que desarrolla el título tercero de la Ley 25/1990, de 20 de diciembre, del Medicamento.

Teniendo en cuenta que lo que desde el punto de vista científico es incorrecto no puede ser aceptable desde el punto de vista ético, si todos

los ensayos clínicos publicados hubiesen sido aprobados por un Comité de Ensayos Clínicos (CEC) no presentarían deficiencias metodológicas importantes, al menos teóricamente. En este sentido, la aprobación por un CEC de un protocolo se consideraría una garantía de calidad desde el punto de vista científico, además de proteger y salvaguardar los derechos humanos de los sujetos participantes. Es necesario tomar conciencia de la importancia que tiene el buen funcionamiento de los actualmente denominados Comités Éticos de Investigación Clínica para mejorar no sólo la calidad de la investigación, sino también de la asistencia sanitaria al poder ofrecer a los sujetos de la investigación alternativas terapéuticas aún no disponibles para la población general.

Conclusiones

Es importante destacar que en algunos casos ciertas deficiencias descritas son tan importantes que ponen en entredicho tanto la validez interna o certeza de los resultados (aleatorización, enmascaramiento y tamaño de la muestra) como la validez externa o capacidad de generalización (pérdidas pre y postaleatorización). Ello lleva a cuestionar la posible efectividad o relevancia clínica de los resultados¹⁸ haciendo que en muchos casos constituyan una mala inversión y en ocasiones induzcan a mala práctica. En otras ocasiones, como la falta de descripción de los acontecimientos adversos, se impide realizar la valoración de la relación riesgo/beneficio imprescindible en terapéutica, y la comparación con otras alternativas disponibles. En fin, los aspectos éticos son fundamentales puesto que estamos investigando en seres humanos y debemos respetar los principios básicos ya enunciados en el informe Belmont en 1978¹⁹ principalmente los de no maleficencia y autonomía que son los más afectados en los casos descritos²⁰.

Con el objeto de mejorar la calidad de los ensayos clínicos, algunos autores²¹ han sugerido medidas a diferentes niveles. Unas a nivel de política editorial como la exigencia de completar por los autores la información proporcionada en unas listas-guía para aceptar su manuscrito, o para los propios revisores (*referees*) que deben emitir un juicio sobre la posible publicación. Otras recomendaciones se dirigen a los posibles autores de manuscritos, resaltando los elementos que suelen olvidarse de incluir en sus artículos, sobre todo en el apartado de “Pacientes y método”. Recientemente, la aparición de revistas electrónicas dan la opción de publicar en papel un resumen estructurado y la posibilidad de conseguir el in-

forme completo de los resultados en la edición electrónica. Esta alternativa hace que los directores de las revistas científicas no puedan aducir la falta de espacio para la reproducción total de los resultados de un ensayo clínico³.

En nuestra opinión es una tarea difícil, en la que a veces pueden aparecer intereses encontrados, pero merece la pena un esfuerzo de todos para mejorar la calidad de los ensayos clínicos en todas sus fases: planteamiento, diseño, realización y publicación de sus resultados.

BIBLIOGRAFÍA

- Chalmers I, Adams M, Dickersin K, Hetherington J, Tarnow-Mordi W, Meinert C et al. A cohort study of controlled trials published as short reports. *JAMA* 1990; 263: 1.401-1.405.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-872.
- Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990; 263: 1.405-1.408.
- Soto J, Galende I, Sacristán JA. Calidad de los ensayos clínicos publicados en España: valoración a través del análisis de tres revistas durante el período 1985-1991. *Med Clin (Barc)* 1994; 102: 241-245.
- Sacristán JA, Soto J, Galende I. Evaluación crítica de ensayos clínicos. *Med Clin (Barc)* 1993; 100: 780-787.
- Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-153.
- Gotzsche PC. Methodology and overt and hidden bias reports of 196 double-blind trials of non-steroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989; 10: 31-56.
- Altman DG. Randomisation, essential for reducing bias. *Br Med J* 1991; 302: 1.481-1.482.
- Grupo de trabajo sobre terminología. Coordinadora: Inés Galende. Glosario de términos de ensayos clínicos. En: García Alonso F, Bakke OM, editores. Metodología del ensayo clínico [monografías del Dr. Esteve, n.º 11]. Barcelona: Fundación Dr. Antonio Esteve, 1991; 113-116.
- Meinert CI. *Clinical Trials: design, conduct and analysis*. Nueva York: Oxford University Press, 1986.
- Galende I, Tristán C. Problemas prácticos en un ensayo clínico (II). En: García Alonso F, Bakke OM, editores. Metodología del ensayo clínico [monografías del Dr. Esteve, n.º 11]. Barcelona: Fundación Dr. Antonio Esteve, 1991; 21-30.
- CPMP Working Party on Efficacy of Medicinal Products. *Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products*. Final, mayo 1993.
- Pocock SJ. *Clinical Trials: a practical approach*. Nueva York: John Wiley, 1983.
- Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979; 300: 1.242-1.245.
- Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *Br Med J* 1984; 289: 1.281-1.284.
- Anónimo. Drop-outs from clinical trials. *Lancet* 1987; 2: 892.
- Real Decreto 944/78, de 14 de abril, por el que se regulan los ensayos clínicos de productos farmacéuticos y preparados medicinales. B.O.E. de 6 de mayo de 1978.
- Wermeling DP, Selwitz AS. Current issues surrounding women and minorities in drug trials. *Ann Pharmacother* 1993; 27: 904-911.
- Informe Belmont. Principios éticos y recomendaciones para la protección de las personas objeto de experimentación. Comisión Nacional para la protección de las personas objeto de la experimentación biomédica y conductual. En: Ensayos clínicos en España (1982-1988). Monografía técnica n.º 17. Madrid: Ministerio de Sanidad y Consumo, 1990; 121-
- Galende I. Problemas éticos de la utilización de seres humanos en la investigación biomédica. Consecuencias sobre la prescripción de medicamentos. *Med Clin (Barc)* 1993; 101: 20-23.
- Spilker B. Improving the standards of clinical trials and clinical trials publications. En: Ruiz Ferrán J, Lahuerta Dal Ré J, Lardinois R, editores. *Communication in Pharmaceutical Medicine: a challenge for 1992*. 7ICPM. Barcelona: Prous Science Publishers, 1991; 3-5.