



Ediciones Doyma, S.L. detenta la propiedad del *copyright* de los artículos incluidos en la presente monografía. Dichos artículos han sido reproducidos tanto en formato impreso como electrónico con permiso de Ediciones Doyma S.L.

29

MONOGRAFÍAS DR. ANTONIO ESTEVE

# ASPECTOS CONCEPTUALES DEL ENSAYO CLÍNICO

Una revisión a través de artículos publicados  
en Medicina Clínica (1990-1999)

F. Bosch

© 2000, Fundación Dr. Antonio Esteve  
Llobet i Vall-Llosera, 2. E-08032 Barcelona.  
Teléfono: 93 433 53 20 Fax: 93 450 48 99.  
Correo electrónico: fundacion@esteve.org  
Internet: <http://www.esteve.org>

Ediciones Doyma S.L. detenta la propiedad del *copyright* de cada uno de los artículos incluidos en la presente monografía. Dichos artículos han sido reproducidos con permiso de Ediciones Doyma S.L.

Depósito legal: B.- 40.878-2000  
Coordinación y producción:  
Ediciones Doyma, S.L.  
Travesera de Gracia, 17-21. E-08021 Barcelona  
Impreso en España por Sorpama, S.A.  
Printed in Spain

*La Fundación Dr. Antonio Esteve contempla como objetivo prioritario el estímulo del progreso de la terapéutica por medio de la comunicación y la discusión científica.*

*La Fundación quiere promover la cooperación internacional en la investigación farmacoterapéutica y, a tal fin, organiza reuniones internacionales multidisciplinarias donde grupos reducidos de investigadores discuten los resultados de sus trabajos. Estas discusiones son recogidas en las publicaciones de los Esteve Foundation Symposia.*

*Otras actividades de la Fundación Dr. Antonio Esteve incluyen la organización de reuniones dedicadas a la discusión de problemas de alcance más local y publicadas en el formato de la presente monografía. La Fundación participa también en conferencias, seminarios, cursos y otras formas de apoyo a las ciencias médicas, farmacéuticas y biológicas y, con carácter bienal, concede un premio al mejor artículo publicado por un autor español dentro del área de la farmacoterapia.*

*Más recientemente, se ha iniciado la publicación de la serie Pharmacotherapy revisited: An Esteve Foundation Series, en la cual a través de diferentes volúmenes se recopilan, en edición facsímil, los principales artículos que sentaron las bases de una determinada disciplina.*

Tanto la introducción como los artículos que recopila la presente monografía recogen la opinión de los correspondientes autores, por lo que la Fundación Dr. Antonio Esteve no se hace necesariamente partícipe de su contenido. Puesto que las filiaciones de los autores corresponden a las que figuran en los artículos originales, no siempre coincidirán con sus filiaciones actuales.

## **AGRADECIMIENTOS**

---

La Fundación Dr. Antonio Esteve desea expresar su agradecimiento a Ediciones Doyma S.L. por autorizar la reproducción de los artículos que aparecen en la monografía. Así mismo agradece a la Dra. Elena Guardiola su colaboración y asesoramiento en las búsquedas bibliográficas efectuadas para la selección de los artículos finalmente incluidos.

---

# Prólogo

---

La sistematización del ensayo clínico como instrumento de investigación ha supuesto una pieza clave en el avance de la medicina en los últimos cincuenta años. Las publicaciones de Bradford Hill, en Gran Bretaña, y de Harry Gold y Louis Lasagna, en Estados Unidos, a mediados del siglo xx, recogen todo un corpus doctrinal sobre la metodología, posibilidades, problemas y limitaciones del ensayo clínico que puede considerarse perfectamente válido e incluso actual. No obstante, la riqueza de este instrumento de conocimiento ha hecho que numerosos autores se preocuparan del mismo y contribuyeran, de un modo u otro, a incrementar su consistencia y a facilitar su difusión.

*Medicina Clínica* es una revista que, desde sus inicios, ha acogido en sus páginas los más relevantes aspectos del conjunto de disciplinas que configuran su nombre. Por ello, no ha de extrañar que resulte una atalaya excepcional desde la

que contemplar el calado que han tenido los problemas relacionados con los ensayos clínicos en la medicina española. La recopilación que se ofrece en esta monografía es, pues, una visión del interés por los ensayos clínicos en España, a través de una variedad de artículos sobre los mismos publicados en *Medicina Clínica*. Es posible que algunos discrepen con determinados aspectos de la selección efectuada por el editor de la monografía, pero estamos seguros de que todos celebrarán el homenaje que esta nueva colaboración entre Ediciones Doyma y la Fundación Dr. Antonio Esteve dedica al interés de los investigadores españoles por los ensayos clínicos, interés que es prueba evidente de la pujanza de la investigación clínica en nuestro país.

**Miquel Vilardell\* y Sergio Erill\*\***

*\* Medicina Clínica.*

*\*\* Fundación Dr. Antonio Esteve.*

---

# Aspectos conceptuales del ensayo clínico

---

---

F. BOSCH	
Introducción	IX

---

## 1 Concepto y diseño

---

J. GARCÍA PUIG, F. MATEOS ANTÓN, A. GIL AGUADO Y C. BARCINA SÁNCHEZ	
¿Puede conocerse el “doble ciego” antes de desvelar los códigos de identificación?	3
M.C. CARRÉ LLOPIS Y J. JIMÉNEZ VILLA	
La investigación clínica de medicamentos en el ámbito extrahospitalario	6
J.B. CABELLO LÓPEZ, V. ABRAIRA Y J. GÓMEZ GARCÍA	
El ensayo clínico para un solo paciente. Justificación, metodología y aportaciones bioéticas	8
F. GARCÍA-ALONSO, E. GUALLAR, O.M. BAKKE Y X. CARNÉ	
El placebo en ensayos clínicos con medicamentos	19
J. SOTO	
Utilidad de la incorporación de análisis de evaluación económica en los protocolos de los ensayos clínicos	26
R. DAL-RÉ, R. ORTEGA Y J. ESPADA	
Eficiencia de los investigadores en la selección de los enfermos para ensayos clínicos: a propósito de un estudio multinacional	28

---

---

X. BADIA Y X. CARNÉ	
La evaluación de la calidad de vida en el contexto del ensayo clínico	33

---

## 2 Bioestadística

---

M.C. CARRÉ LLOPIS, J. JIMÉNEZ VILLA, M. MARTÍN MATEO Y F. JANÉ CARREÑA	
La estadística en la investigación clínica de medicamentos. Estudio de artículos originales procedentes de centros españoles	47
C. RUBIO TERRÉS	
Diseño estadístico de ensayos clínicos	56
E. GUALLAR, F.J. JIMÉNEZ, F. GARCÍA-ALONSO Y O.M. BAKKE	
La regresión a la media en la investigación y práctica clínica	66
A. CASADO, L. PRIETO Y J. ALONSO	
El tamaño del efecto de la diferencia entre dos medias: ¿estadísticamente significativo o clínicamente relevante?	72

---

## 3 Calidad y buenas prácticas clínicas

---

J.A. SACRISTÁN, J. SOTO E I. GALENDE DOMÍNGUEZ	
Evaluación crítica de ensayos clínicos	81
I. GALENDE DOMÍNGUEZ, J.A. SACRISTÁN Y J. SOTO	
Cómo mejorar la calidad de los ensayos clínicos	92

---

Dentro de cada tema, los artículos se han clasificado en orden ascendente según su fecha de publicación en *Medicina Clínica*.

C. VALLVÉ	
La buena práctica clínica y la caja de Pandora (y III). La fiabilidad de los datos	101
G. URRUTIA, X. BONFILL Y J. MARTÍ	
<i>En busca de la evidencia perdida.</i> Propuestas para incrementar la transparencia de los ensayos clínicos	105
F.J. GARCÍA LÓPEZ, S. GUTIÉRREZ BEZÓN, I. GALENDE DOMÍNGUEZ Y C. AVENDAÑO SOLA	
Evaluación de la calidad de los ensayos clínicos: justificación, utilidad e inconvenientes	115

#### 4 Ética médica

J. MORENO Y B. GONZÁLEZ GABALDÓN	
Investigación en humanos	129
D. GRACIA	
Entre el ensayo clínico y la ética clínica: las buenas prácticas clínicas	134
I. GALENDE DOMÍNGUEZ	
Problemas éticos de la utilización de seres humanos en investigación biomédica. Consecuencias sobre la prescripción de medicamentos	138
J. CAMÍ	
Conflicto de intereses e investigación clínica	144

#### 5 Consentimiento informado

R. DAL-RÉ	
Investigación clínica con fármacos: estudio de la influencia de la información sobre reacciones adversas en la obtención del consentimiento informado	155
P. SIMÓN LORDA Y L. CONCEIRO CARRO	
El consentimiento informado: teoría y práctica (I)	161
P. SIMÓN LORDA, I.M. BARRIO CANTALEJO Y L. CONCEIRO CARRO	
Legibilidad de los formularios escritos de consentimiento informado	168

J.P. ORDOVÁS BAINES, E. LÓPEZ BRIZ, E. URBIETA SANZ, R. TORREGROSA SÁNCHEZ Y N.V. JIMÉNEZ TORRES	
Análisis de las hojas de información al paciente para la obtención de su consentimiento informado en ensayos clínicos	179

#### 6 Comités de ética

M.M. LÓPEZ MARTÍN Y J.E. BATISTA-MIRANDA	
Comités éticos de investigación: una complicación necesaria	189
M.I. LUCENA, G. TOGNONI Y F. SÁNCHEZ DE LA CUESTA	
Comités éticos de investigación clínica: nuevas funciones para un nuevo escenario	195

#### 7 Publicaciones biomédicas

B. SPILKER	
Cómo elevar la calidad de los ensayos clínicos y de su publicación	207
S. LOCK	
La revisión de manuscritos	210
T.C. CHALMERS	
La disminución del sesgo en la publicación de los ensayos clínicos	214
S. ERILL	
La ética de la publicación: el caso de los resultados negativos	219
P. BROWN	
Editoriales, revisiones y cartas: por quién, para quién y cuándo	222

#### 8 Revisiones bibliográficas

J.M. LÓPEZ ARRIETA Y N. QIZILBASH	
La medicina basada en pruebas: revisiones sistemáticas. La Colaboración Cochrane	227
J. MARTÍ, X. BONFILL, G. URRUTIA, J.R. LACALLE Y R. BRAVO	
Identificación y descripción de los ensayos clínicos publicados en revistas españolas de medicina general e interna durante el período 1971-1995	235
Índice cronológico	247



---

# Introducción

---

Cuando en todos los ámbitos, científicos y no científicos, se están planteando las perspectivas de futuro del nuevo milenio, concretamente en el campo del ensayo clínico (EC) nos hallamos en un momento, a mi entender, al menos etiquetable de significativo. El por qué, se puede razonar, desde una visión muy personal, desde tres puntos de vista: el pasado, el presente y el futuro. En cuanto al primero, recientemente se celebró el cincuentenario del estudio que sentó las bases del EC: auspiciado por el *Medical Research Council*, diseñado por Sir Austin Bradford Hill (1897-1991) y publicado en 1948. Aunque históricamente se deberían citar también otros trabajos de relevancia que aportaron importantes avances en el diseño y la metodología del EC, se acepta que fue el trabajo de Hill y su personal dedicación e insistencia por establecer las bases científicas para discernir la eficacia de nuevas intervenciones terapéuticas, que propiciaron el nacimiento del EC comparativo. A pesar de la resistencia del colectivo médico de la época, Hill consiguió que la aleatorización en la asignación de los grupos comparativos de los EC fuera un procedimiento sinónimo de calidad, que con el tiempo se convertiría en prácticamente indispensable. El presente del EC podría definirse, al menos en nuestro país, como la adquisición y maduración de técnicas y conocimientos anteriormente ya implantados en otros países más avanzados. A raíz de la Ley del Medicamento de 1990 se orientó la investigación de los medicamentos hacia unos cauces mucho más estandarizados comparativamente con épocas precedentes y con el Real Decreto 561 de 1993, específicamente sobre los EC con medicamentos, se sentaron las bases que permiten hablar de una situación al menos regularizada en España. El futuro conlleva a plantearse todavía muchas asignaturas pendientes, entre las cuales cabría citar la necesidad de una mejor información del medicamento tanto hacia los partici-

pantes de los EC, como hacia los profesionales sanitarios y el ciudadano en general. Además, los conflictivos aspectos éticos, la confidencialidad, la correcta metodología y la evaluación de los resultados son otros temas que, junto con los problemas de ámbito regulador, merecerían especial atención en el futuro. La agilización de toda la gestión que envuelve la realización del EC, la formación en investigación clínica de los diferentes profesionales sanitarios y la potenciación de la aplicación de estas técnicas en intervenciones no farmacológicas, junto con la promoción de EC desde instituciones académicas, autoridades sanitarias y centros de investigación independientes, podrían ser también otros objetivos que contribuirían a un más completo desarrollo de la investigación clínica.

Partiendo de la idea de que *Medicina Clínica* constituye una revista de referencia en el panorama de la investigación clínica española, se pensó que las publicaciones sobre aspectos conceptuales de los EC aparecidas en esta revista durante la última década, podrían aportar una visión suficientemente amplia de la situación de los EC en nuestro país. La inevitable limitación de espacio que implica pretender encajar esta idea en el formato de la serie Monografías Dr. Antonio Esteve, condicionó el tener que seleccionar, por un proceso de marcada subjetividad, un número muy limitado de publicaciones.

Si bien se partió de la base bibliográfica en red PubMed recuperando aquellos artículos indizados con las palabras claves "ensayo/s clínico/s" (*clinical trial/s*) atribuidas a *Medicina Clínica* entre 1990 y 1999, se recurrió asimismo a búsquedas en sistemas CD-ROM (Medline y *Medicina Clínica*) y a búsquedas manuales a través de la versión impresa de la revista. Para completar la obtención de un archivo general de partida con más de 360 artículos sobre EC se incluyeron publicaciones recuperadas a partir de otras palabras clave relacionadas (p. ej., metaanálisis, evaluación o investigación

de fármacos, aleatorización, doble ciego, estudios controlados o multicéntricos, ética, consentimiento informado, legislación, etc.). Se descartaron de entrada los artículos sobre EC de evaluación de intervenciones específicas o los que se referían a aspectos muy concretos de una determinada enfermedad y se prefirió seleccionar aquellos más generales, que abarcaran principalmente aspectos conceptuales o que invitaran a la reflexión o la polémica. Aparte de una selección por temas, evitando excesivas repeticiones o solapamientos, en la medida de lo posible se procuró mantener la representatividad de los autores en el sentido de ser más estrictos al decidir la inclusión de uno de ellos en más de una ocasión. Al mismo tiempo, se pretendió conseguir una cierta representatividad geográfica en la procedencia de los trabajos, así como limitar la inclusión de un número excesivo de estudios que se hubiesen publicado en un mismo año. Se excluyeron los artículos sobre legislación y las cartas al director u otros comentarios, con la excepción de que el tema tuviera interés especial y no hubiese sido tratado en otra de las publicacio-

nes seleccionadas. Después de este arduo e, in-sisto, subjetivo proceso se obtuvieron las treinta y tres publicaciones (expurgadas de resúmenes en español e inglés) que definitivamente se incluyen en el presente volumen. No se ha pretendido, ni mucho menos, realizar una selección de calidad puesto que todos los artículos por el hecho de haber sido aceptados en una revista con revisores especializados se entiende que cumplen unos mínimos estándares de calidad.

Merecen una mención especial los diferentes autores que con sus trabajos han contribuido a recoger un amplio abanico de aspectos de interés sobre los EC y sus principales controversias. Por último, y como principal objetivo, con esta nueva colaboración entre *Medicina Clínica* y la Fundación Dr. Antonio Esteve se pretende ofrecer una obra de utilidad tanto para los estudiantes como para los diferentes profesionales involucrados en el mundo de la investigación clínica.

**Félix Bosch**

Fundación Dr. Antonio Esteve

Parte 1

---

**CONCEPTO Y DISEÑO**

---

---

# ¿Puede conocerse el “doble ciego” antes de desvelar los códigos de identificación?

---

Juan García Puig, Felícitas Mateos Antón<sup>a</sup>, Antonio Gil Aguado y Carlos Barcina Sánchez<sup>b</sup>

Servicios de Medicina Interna y <sup>a</sup>Bioquímica Clínica. Hospital La Paz. <sup>b</sup>Departamento Médico. Laboratorios Zeneca Farma, S.A. Madrid.

El término “ciego” se refiere a un procedimiento de investigación clínica que oculta la identidad de un fármaco o procedimiento terapéutico. Se utiliza para reducir posibles sesgos en la evaluación terapéutica y evitar el efecto placebo que con frecuencia se produce en los ensayos abiertos. “Doble ciego” significa que tanto el médico como el paciente ignoran la identidad del fármaco o procedimiento terapéutico que se pretende examinar. En ciertos ensayos clínicos (EC) la información acumulada puede desvelar el método “doble ciego” (resultados de laboratorio, reacciones adversas concretas), de forma que los resultados finales podrían estar influenciados por este hecho<sup>1</sup>. Al finalizar un EC en el que se compararon 2 fármacos antihipertensivos, hemos evaluado el grado de concordancia entre la asignación “ciega” realizada por los investigadores y la asignación real según un método aleatorio.

Hemos participado en un EC multicéntrico, internacional, denominado BRILLIANT (Blood pressure, Renal effects, Insulin control, Lipids, Lisinopril And Nifedipine Trial) que pretendía comparar el efecto de lisinopril con el del nifedipino *retard* durante 12 meses en la excreción renal de albúmina en pacientes hipertensos con diabetes mellitus tipo II y microalbuminuria (entre 20 y 200 µg/min). El EC se realizó de acuerdo con las normas de buena práctica clínica<sup>2</sup> y los fármacos se administraron según un procedimiento “doble ciego” con doble enmascaramiento. Durante el estudio se determinó la excreción de microalbuminuria en situación basal y a los 6 y 12 meses, mediante la recogida de 3 muestras de orina nocturna de 8 h. Todas las muestras de orina se recogieron en los centros participantes y se enviaron a un laboratorio de referencia en Inglaterra. Los resultados se comunicaron a los investigadores por telefax y correo en un plazo inferior a

15 días, tras la recepción de las muestras urinarias. En cada visita clínica el paciente debía contestar a la siguiente pregunta: “¿Ha tenido usted algún problema de salud desde la última visita?”. Al finalizar el estudio, y antes de conocer la asignación real aleatoria establecida por el promotor, efectuamos una asignación “ciega” de acuerdo con la comunicación voluntaria de reacciones adversas por los pacientes y a las modificaciones de la excreción renal de albúmina. Para esta asignación “ciega” empleamos los siguientes criterios:

*Asignación a nifedipino.* Pacientes que durante el tratamiento activo refiriesen algún efecto secundario atribuible a nifedipino de grado posible o probable, tales como enrojecimiento facial, sensación de calor, cefalea persistente o edemas en miembros inferiores. También se consideró que los pacientes estaban tomando nifedipino cuando, en ausencia de efectos secundarios atribuibles a la administración de nifedipino, la microalbuminuria no descendió más de un 20 % a los 6 o 12 meses, con respecto a la cifra mediana basal de tres determinaciones consecutivas en la orina nocturna de 8 h.

*Asignación a lisinopril.* Pacientes que a lo largo del estudio refiriesen efectos secundarios atribuibles a lisinopril de grado posible o probable, como molestias de garganta, carraspera o tos persistente. También se consideró que los pacientes estaban tomando lisinopril cuando, en ausencia de efectos secundarios referidos por el paciente, la microalbuminuria descendió más de un 20 % a los 6 o 12 meses, con respecto a la cifra mediana basal de tres determinaciones consecutivas en la orina nocturna de 8 h.

La asignación “ciega” se realizó por cada uno de los firmantes antes de conocer la asignación real aleatoria. Se dio prioridad a la aparición de efectos secundarios atribuibles a la medicación

TABLA I  
 CONCORDANCIA ENTRE LA ASIGNACIÓN REAL ALEATORIA Y LA ASIGNACIÓN "CIEGA"  
 ANTES DE ABRIR LOS CÓDIGOS DE IDENTIFICACIÓN, DE 4 INVESTIGADORES  
 PARTICIPANTES EN UN ENSAYO CLÍNICO QUE COMPARÓ LA EFICACIA Y SEGURIDAD DE  
 LISINAPRIL FRENTE A NIFEDIPINO SOBRE LA EXCRECIÓN RENAL DE MICROALBUMINURIA  
 EN PACIENTES HIPERTENSOS CON DIABETES MELLITUS TIPO II

	Investigador 1		Investigador 2		Investigador 3		Investigador 4	
	L	N	L	N	L	N	L	N
Asignación L	12	2	13	2	13	2	14	2
"ciega" N	4	15	3	15	3	15	2	15

El índice kappa menor ( $0,635 \pm 0,134$ ; intervalo de confianza del 95 %,  $0,372-0,898$ ;  $p < 0,05$ ) fue obtenido por el investigador número 1 (27 de 33 posibles aciertos); L: lisinapril; N: nifedipino.

sobre las variaciones de la microalbuminuria. No tuvimos en cuenta las variaciones tensionales por considerar que el efecto hipotensor de ambos fármacos sería de magnitud semejante.

Después de 4 semanas con placebo (simple ciego) un total de 36 enfermos (22 varones, 14 mujeres; edad media, 61 años) fueron aleatorizados a recibir lisinapril o nifedipino. Al finalizar los 12 meses de tratamiento activo, en 3 enfermos no se pudo efectuar una asignación "ciega" por abandono precoz del estudio sin haber presentado efectos secundarios atribuibles a los fármacos del ensayo (muestra global, 33 enfermos). Cuando el promotor nos proporcionó el código de aleatorización (asignación real), pudimos comparar estos resultados con los obtenidos mediante la asignación "ciega". Los resultados de los 4 firmantes se recogen en la tabla I. De los 16 pacientes tratados con lisinapril (asignación real), hubo concordancia con la asignación "ciega" en 12-14 casos (75,0-87,5 %). Nosotros estimamos que 12-14 enfermos estaban tomando lisinapril: en 13 pacientes, sin efectos secundarios relevantes, se evidenció un marcado descenso de la microalbuminuria (valor basal medio  $85 \mu\text{g}/\text{min}$ ; valor final medio,  $11 \mu\text{g}/\text{min}$ ;  $p < 0,01$ ). Un paciente presentó tos persistente que obligó a suspender la medicación antes de los 6 meses, y en otro coincidió la presencia de tos y descenso de microalbuminuria (basal,  $77 \mu\text{g}/\text{min}$ ; 12 meses,  $15 \mu\text{g}/\text{min}$ ). De los 17 enfermos tratados con nifedipino (asignación real), en 15 la asignación "ciega" fue correcta (88 %) y postulamos que otros 2-4 pacientes también habrían recibido nifedipino. Catorce enfermos refirieron efectos secundarios de posible o probable relación con nifedipino (edemas en miembros inferiores, 11 enfermos; cefalea tensional, 2 enfermos; enrojecimiento facial, un enfermo); y en 3-5 pa-

cientes, sin efectos secundarios relacionables con nifedipino, hubo un incremento de la excreción de albúmina (valor basal medio  $57 \mu\text{g}/\text{min}$ ; valor final medio,  $98 \mu\text{g}/\text{min}$ ;  $p < 0,05$ ). El índice kappa global observado (concordancia entre la asignación "ciega" y la asignación real) fue de  $0,635 \pm 0,134$  (intervalo de confianza del 95 %,  $0,372-0,898$ ;  $p < 0,05$ ), lo que indica una concordancia aceptable<sup>3</sup>.

Estos resultados indican que el método "doble ciego" puede ser desvelado por informaciones propias del EC: en 27-29 de 33 enfermos (81,8-87,9 %) la asignación "ciega" a lisinapril o nifedipino coincidió con la asignación real aleatoria. Esta concordancia se basó en el diferente espectro de efectos secundarios de lisinapril y nifedipino<sup>4,5</sup> y en la diferente acción de ambos fármacos sobre la excreción renal de albúmina<sup>6-8</sup>.

De estos resultados surgen dos interrogantes: a) ¿ha podido influir en los resultados globales del EC el elevado tanto por ciento de aciertos de la asignación "ciega"? y b) ¿cómo se puede mejorar el método "doble ciego"? La asignación "ciega" no ha podido determinar los resultados globales del estudio multicéntrico por diversas razones: en el estudio se han incluido un total de 335 pacientes de forma que nuestra aportación al conjunto del estudio ha sido próxima al 10 %. En muchos casos los efectos adversos aparecieron tardíamente y los resultados de la microalbuminuria se conocieron a los 6 y 12 meses de la aleatorización. Los 4 firmantes de esta carta, dos de los cuales no han desempeñado funciones clínicas en el estudio, alcanzamos un tanto por ciento de concordancia similar entre la asignación real aleatoria y la asignación "ciega" (81,8, 85,3, 84,8 y 87,9 %, respectivamente; media, 85,0 %).

El método "doble ciego", habitualmente utilizado en los EC, puede ser perfeccionado me-

dante la ocultación de datos no relevantes para el cuidado del enfermo<sup>1</sup>. En este EC los valores de la microalbuminuria podían haber sido comunicados al equipo investigador cuando el estudio hubiese finalizado en todos los centros participantes. El "doble ciego" también puede mejorarse con la participación de 2 observadores en la evaluación clínica y desarrollo del estudio<sup>1</sup>, de forma que uno de ellos no pueda tener acceso a información reveladora del fármaco administrado antes de la apertura del código de asignación aleatoria.

## BIBLIOGRAFÍA

1. Spilker B. Choosing and validating the clinical trial's blind. En: Spilker B, editor. Guide to clinical trials. Nueva York: Raven Press, 1991; 15-20.
2. Real Decreto 561/1993, de 16 de Abril, por el que se establecen los requisitos para la realización de ensayos clínicos con medicamentos. BOE 13/5/93.
3. Fleiss JL. Statistical methods in biological assay (3.ª ed.). Nueva York: McMillan, 1991.
4. Marley JE. Safety and efficacy of nifedipine 20 mg tablets in hypertension using electronic data collection in general practice. J Roy Soc Med 1989; 82: 272-275.
5. Morales-Olivas FJ, Ferrer JM, Palop V, Rubio E. Reacciones adversas por fármacos inhibidores de la enzima conversiva de la angiotensina notificadas por medio de tarjeta amarilla. Med Clin (Barc) 1994; 103: 321-325.
6. Ranieri G, Filitti V, Bonfantino MV, De Cesaris R. Effects of lisinopril and amlodipine on microalbuminuria in hypertensive patients [resumen]. Am J Hypertens 1992; 5: 113.
7. Bianchi S, Bigazzi R, Baldari G, Campese VM. Microalbuminuria in patients with essential hypertension: effects of several antihypertensive drugs. Am J Med 1992; 93: 525-528.
8. Puig JG, Mateos FA, Ramos TH, Lavilla MP, Capitán MC, Gil A. Albumin excretion rate and metabolic modifications in patients with essential hypertension: Effects of two angiotensin converting enzyme inhibitors. Am J Hypertens 1994; 7: 46-51.

---

# La investigación clínica de medicamentos en el ámbito extrahospitalario

---

María del Carmen Carré Llopis<sup>a</sup> y José Jiménez Villa<sup>b</sup>

<sup>a</sup>Grupo de Docencia e Investigación en Salud (GDIS). Barcelona. <sup>b</sup>Regió Sanitària Costa de Ponent. Servei Català de la Salut. L'Hospitalet de Llobregat. Barcelona.

La investigación clínica de medicamentos se realiza actualmente casi en exclusiva en el mundo hospitalario, incluso cuando se estudian tratamientos e indicaciones que se usan en el ámbito extrahospitalario. De los 288 artículos de investigación clínica de medicamentos realizados en centros españoles e indizados en la base de datos de Excerpta Médica en los años 1980, 1985 y 1990, únicamente seis (2,08%) declararon haberse realizado en el ámbito extrahospitalario y todos ellos fueron publicados en revistas nacionales.

La mayoría de artículos sobre ensayos clínicos comparativos de medicamentos en fase IIIb o incluso IV llevados a cabo en centros hospitalarios se han realizado con muestras homogéneas de pacientes seleccionados por criterios estrictos, muy distintos de los que se visitan en el medio extrahospitalario. Si bien este tipo de estudios es el más indicado para determinar la eficacia de un medicamento en situaciones "ideales", sus resultados sólo pueden extrapolarse a poblaciones muy concretas y seleccionadas, seguidas y evaluadas por unos criterios estrictos y con medios diagnósticos normalmente sofisticados que hacen que la eficacia observada no se corresponda con la efectividad que se observará al utilizarlo en la práctica extrahospitalaria. Por el contrario, el ámbito extrahospitalario precisa de estudios más orientados a evaluar la efectividad de los tratamientos en la práctica diaria, pues su interés radica en alcanzar el máximo impacto en la población. Este tipo de estudios presentan algunos elementos más específicos, intentando reproducir las características de la práctica habitual, estudiando muestras heterogéneas más representativas de la población demandante<sup>1,2</sup>. En el ámbito de la atención primaria de la salud ya empiezan a observarse estudios de este tipo<sup>3</sup>, aun-

que todavía son infrecuentes en otras especialidades médicas extrahospitalarias.

La realización de estudios con muestras heterogéneas de pacientes obliga a estudiar un mayor número de sujetos para poder extraer conclusiones fiables, por lo que difícilmente pueden ser llevados a cabo en un solo centro. Sin embargo, ello no implica que no deban realizarse con el mismo rigor científico, sino que incluso deben ser diseñados y monitorizados con mayor cuidado ya que son muchos los puntos del estudio en los que podrían introducirse sesgos.

Ya en los años sesenta Schwartz y Lellouch<sup>2</sup> definieron dos tipos de ensayos claramente diferenciados, los que denominaron *ensayos explicativos*, que son los que ofrecen el máximo control de todos los aspectos del experimento, estudiando muestras de sujetos altamente seleccionadas y criterios de seguimiento y evaluación muy estrictos y los *ensayos pragmáticos* que intentan estudiar muestras de sujetos más representativos, con criterios de seguimiento y evaluación similares a los que se utilizaron en el ámbito de la práctica extrahospitalaria, pero que no por ello están diseñados y llevados a cabo de forma menos rigurosa. Estos últimos son los más adecuados para llevar a cabo estudios extrahospitalarios<sup>4</sup>, ya que estudian la efectividad de los tratamientos en las indicaciones y condiciones reales de uso, o sea en pacientes que presentan diversas enfermedades, habitualmente crónicas de grado leve o moderado o incluso pacientes bien controlados. Además las consultas del medio extrahospitalario disponen de recursos diagnósticos y de exploración limitados.

En España no existe tradición de realizar estudios multicéntricos<sup>5</sup>. Además, debido a la estructura sanitaria de nuestro país la investigación extrahospitalaria presenta con frecuencia una serie de inconvenientes como la falta de disponibilidad de tiempo, de recursos y de organización en equipos multidisciplinarios que puedan

llevar a cabo estudios multicéntricos, y su mayor presión asistencial. Aunque recientemente se están realizando algunas experiencias puntuales, sería conveniente que, desde las diversas instituciones y sociedades científicas, se promoviera la creación de grupos de trabajo multidisciplinarios, que diseñen y realicen de forma adecuada estudios de la evaluación de la efectividad y seguridad de medicamentos en los pacientes y condiciones en que suelen atenderse en el ámbito extrahospitalario.

El diseño del protocolo de estudio debe estar basado en criterios realistas y factibles, a la vez que homogéneos, aceptados y seguidos rigurosamente por todos los médicos investigadores. Por ello, los grupos de trabajo deberían ser liderados y coordinados por personal competente y eficiente, capaz de impulsar la investigación y de mantener el interés a lo largo de todo el estudio. Sería conveniente que todos los investigadores tuvieran la oportunidad de, a partir de versiones preliminares del protocolo de estudio, acordar los criterios de selección, exploración, seguimiento y evaluación según las necesidades y disponibilidades de su propia práctica. Posteriormente, se deberían establecer un calendario y unas normas

para asegurar la homogeneidad de criterios, seguimiento y la forma de cumplimentación de los cuadernos de recogida de datos.

Si estos estudios se diseñan y realizan de forma rigurosa, tendrán gran impacto sobre la práctica médica habitual en pacientes ambulatorios, que no olvidemos suponen la mayoría de consultas que se realizan en nuestro país.

## BIBLIOGRAFÍA

1. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; 301: 1.410-1.412.
2. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis* 1967; 20: 637-648.
3. Jiménez Villa J, Carré Llopis MC, Argimón Pallás JM. Tipos de estudios publicados en la revista *Atención Primaria* (1984-1990). *Aten Primaria* 1993; 12: 9-14.
4. Argimón Pallás JM, Jiménez Villa J. *Métodos de investigación aplicados a la atención primaria de la salud*. Barcelona: Doyma, 1991.
5. Gascón ME, Muro P, Nolasco A, Álvarez-Dardet C, Bolumar F. La investigación clínica en España: características de los autores. *Med Clin (Barc)* 1986; 87: 444-447.



---

# El ensayo clínico para un solo paciente. Justificación, metodología y aportaciones bioéticas

---

Juan B. Cabello López<sup>a</sup>, Víctor Abaira<sup>b</sup> y Javier Gómez García<sup>a,c</sup>

<sup>a</sup>Unidad de Investigación. Hospital General Universitario de Alicante. <sup>b</sup>Unidad de Bioestadística Clínica. Hospital Ramón y Cajal. Madrid. Centro Coordinador de REUNI. Instituto de Salud Carlos III. Madrid. <sup>c</sup>Becario del Instituto de Salud Carlos III n.º 96/4448.

La actividad clínica constituye un proceso de decisión-acción relativo a un paciente individual y con una finalidad multidimensional. En efecto, la conducción de la anamnesis y la realización del examen y de unas determinadas pruebas diagnósticas (y no otras), así como la selección de un tratamiento, implican continuas decisiones. La ejecución activa de algunas pruebas (p. ej., pruebas hormonales) o la administración del tratamiento son, por su parte, acciones con diferentes niveles de complejidad y diferentes actores. En unos casos la acción y su control son fácilmente realizables, son los casos de acciones de ejecución directa, por ejemplo realizar una angioplastia, una intervención quirúrgica o administrar una inyección intravenosa; tanto la acción como su control (prescripción y ejecución) son realizables por el médico. En otros casos, por ejemplo, la ejecución de un tratamiento con bloqueadores, amitriptilina o teofilina, la realización de la acción y su control ofrece más dificultades por prolongarse en el tiempo y sobre todo por depender también del paciente; así pues, prescripción y observancia son realizadas, respectivamente, por médico y paciente<sup>1</sup>.

Por otra parte, los objetivos de un tratamiento pueden ser diversos y afectar a múltiples niveles: mejorar los síntomas de la enfermedad, evitar recidivas de ésta, prevenir deterioros o acontecimientos futuros, remediar las molestias actuales o dejar morir con dignidad al paciente<sup>2,3</sup>. La definición de objetivos terapéuticos y la ejecución de los mismos son, con las precisiones pertinentes, un proceso común de médico y de paciente.

Este procedimiento de decidir para un individuo está basado en el uso combinado de evidencias y valores. Respecto a las evidencias, se usa fundamentalmente conocimiento operativo

construido a partir de grupos de pacientes y también conocimiento sobre los mecanismos de enfermar y sanar<sup>4,5</sup>. Como caso particular de este proceso, para la selección de la terapéutica se acepta que el mejor conocimiento sobre tratamientos proviene de los ensayos clínicos controlados y aleatorizados<sup>6</sup>.

Los objetivos del presente artículo son: *a)* plantear los problemas teóricos que existen al aplicar el conocimiento de los ensayos clínicos a casos particulares; *b)* revisar los aspectos más relevantes del diseño, conducción y análisis de los ensayos clínicos para un solo paciente (EC  $n = 1$ ), y *c)* señalar las peculiaridades bioéticas de los EC  $n = 1$ .

## **Problemas al aplicar el conocimiento de los ensayos clínicos a casos particulares**

---

El ensayo clínico es el mejor diseño para conocer la relación entre una acción o tipo de tratamiento y un efecto o resultado del ensayo. Dicho de otro modo, es la mejor forma de obtener evidencias sobre la eficacia, efectividad y seguridad de los procedimientos o fármacos para lograr unos objetivos terapéuticos definidos<sup>6</sup>. Pese a la firmeza epistemológica de este conocimiento, existen algunos aspectos de estos estudios que conviene considerar, por cuanto pueden limitar su valor al seleccionar un tratamiento para un paciente concreto.

Dichos ensayos se realizan, usualmente, en un grupo de pacientes definidos por los criterios de elegibilidad (*población de estudio*), y que serán asignados mediante un procedimiento aleatorio a los grupos a comparar. De este modo, el resultado del estudio se refiere, en sentido estricto, sólo a esa población que reúne los criterios de inclusión. El clínico tendrá que decidir en qué medida su paciente es análogo a la población del estudio y, por tanto, en qué medida cabe esperar en él un comportamiento análogo del tratamiento.

---

Publicado en *Med Clin (Barc)* 1997; 109 (15): 592-598. Este trabajo ha sido financiado por el FIS (proyecto 97/0043-01).

La valoración de la eficacia en un ensayo clínico depende de modo crítico de cuál sea la pregunta principal; en este sentido, los ensayos se diseñan para determinada/s pregunta/s principal/es (habitualmente una), mientras que otras preguntas secundarias influyen menos en el diseño. Por ejemplo, podemos preguntarnos: ¿el fármaco A disminuye la mortalidad o la incidencia de un nuevo accidente coronario?, ambas podrían ser preguntas principales. Las preguntas secundarias podrían ser: ¿el fármaco A mejora el número de crisis de angina, la clase funcional del paciente o la calidad de vida?

Estas preguntas secundarias, aunque suelen estar recogidas en el estudio, no reciben la misma atención en el diseño que las variables principales: por ejemplo, no son sometidas a procesos específicos de mejora de la calidad en la medición (en estudios piloto) y no se evalúan con la misma intensidad que aquéllas. Por otra parte, estas preguntas secundarias no influyen en la estimación del tamaño muestral, y por tanto el diseño del estudio no está condicionado por esas preguntas. Por todo ello, respecto a las preguntas secundarias la calidad de la evidencia no es tan solvente, desde una perspectiva epistemológica, como para las preguntas principales.

En la selección de las preguntas principales influye la percepción de cuál es el objetivo más importante del tratamiento de esa enfermedad, y se eligen a partir de lo que los médicos o investigadores consideran importante para el paciente, por ejemplo prolongar su supervivencia. Sin embargo, el paciente podría tener otros valores y decidir qué otro objetivo es el que a él le importa; esta circunstancia es esencial al elegir el tratamiento y debería considerarse en el diseño de los estudios<sup>7</sup>.

Por otra parte, y salvo que el tratamiento sea extraordinario, en el grupo experimental existirán sujetos en los que el tratamiento estudiado no es efectivo (sujetos no respondedores), ello es así, aun cuando la terapéutica ensayada pueda ser eficaz en el conjunto de pacientes, y la respuesta sea más positiva en este grupo que en el de control. En este sentido, en los ensayos clínicos los indicadores de asociación entre tratamiento y efecto (*odds ratio* [OR], *risk ratio* u otros)<sup>8</sup> informan sobre la intensidad de una relación, de su independencia del azar y de su posible causalidad. Es decir, afirmar que el tratamiento A es 5 veces mejor que el B y que la asociación no depende del azar (OR = 5; *intervalo de confianza* del 95 %, 1,9-9,2) es sólo una *proposición causal potencial* que informa sobre la asociación, pero nada predice sobre un paciente<sup>9</sup>. La proba-

bilidad de que un paciente responda al tratamiento será la proporción de respuestas en el grupo tratado con A o B, y ésta será una *proposición causal predictiva*<sup>9</sup>.

Con esa evidencia de efecto y esa probabilidad de respuesta, indicaremos un tratamiento de acuerdo con el paciente y sus valores; este tratamiento tendrá frecuentemente un sentido de prueba, y por tanto será preciso controlar sus efectos. Puede ocurrir, es lo más probable, que el tratamiento sea eficaz, pero también que no lo sea e incluso que empeore la situación de nuestro paciente o que produzca efectos adversos.

Para algunos objetivos del tratamiento cuya variable de respuesta es la ocurrencia de un acontecimiento a largo plazo, por ejemplo, mortalidad o nuevo accidente coronario, no podremos saber si nuestro tratamiento está siendo efectivo, y sólo cuando el episodio ocurra podremos pronunciarnos retrospectivamente al respecto. Para otros objetivos, cuyas variables de respuesta son índices de estado (como disnea, capacidad funcional, rigidez matutina, dolor articular, etc.) o episodios contabilizables (número de crisis de angina, de episodios de taquicardia o de crisis epilépticas), sí es posible saber si el tratamiento está siendo efectivo o no, e incluso si está perjudicando. No obstante, en la actividad clínica ese pronunciamiento es frecuentemente difícil.

En resumen, existirán pacientes en los que el resultado del ensayo no es aplicable puesto que reúnen criterios de exclusión y no hubieran participado en el estudio. Otras veces, no se ha realizado el ensayo clínico que necesitaríamos para tomar nuestra decisión, o se ha realizado sobre preguntas diferentes de nuestros objetivos de tratamiento. Finalmente puede ocurrir que, aunque el tratamiento fue correctamente seleccionado, existan dudas sobre si está siendo efectivo, sobre la dosis adecuada y sobre si puede estar perjudicando o produciendo efectos adversos.

Cuando concurren algunas de estas circunstancias (tabla I) y se cumplen algunos requisitos de viabilidad (tabla II), la realización de un EC  $n = 1$  es una alternativa a considerar, que suscita cada vez más interés<sup>10,11</sup>.

## Diseño de EC $n = 1$

En esencia, el EC  $n = 1$  es una estrategia estructurada de "prueba de tratamientos" para determinar la efectividad de un tratamiento o identificar efectos adversos en un paciente. Estos estudios proceden de áreas de la psicología y la psiquiatría donde se han propuesto diversos diseños cuya descripción se aleja de los objetivos

TABLA I  
¿CUÁNDO ESTÁ INDICADO HACER  
EC N = 1? (PERTINENCIA)

I. En relación con la efectividad del tratamiento El clínico y/o el paciente dudan de la efectividad de la terapéutica administrada El clínico duda si el tratamiento ha comenzado a actuar en su paciente El paciente insiste en tomar un tratamiento que el clínico considera que no está siendo efectivo
II. En relación con la dosis Existen dudas respecto a la dosis óptima para este paciente
III. En relación con los efectos adversos El paciente experimenta síntomas que pudieran ser atribuibles al tratamiento (pero no es seguro)
IV. El EC n = 1 como test diagnóstico En aquellos test estímulo-respuesta cuya evaluación es clínica y existen dudas sobre la respuesta al mismo en nuestro paciente (p. ej., hidrocortisona en Addison o pidostigmina en miastenia grave)

EC n = 1: ensayos clínicos para un solo paciente.

de este artículo<sup>10-17</sup>. En la actividad clínica el diseño más empleado es el propuesto por Guyatt<sup>10-12</sup>, y a él nos referiremos a partir de aquí.

*Definición*

El EC n = 1 es un estudio (fig. 1) en el cual se prueban en un solo paciente dos tratamientos, un tratamiento en cada subciclo, y se repite la prueba durante varios ciclos. El orden de los tratamientos dentro de cada ciclo es asignado aleatoriamente, y el procedimiento se realiza con doble enmascaramiento de médico y paciente (EC n = 1, ensayo con un solo sujeto, doble enmascaramiento y cruzamiento múltiple).

Desde un punto de vista formal, si se prueba un tratamiento y se observa una mejoría del paciente, ello puede ser debido a la eficacia real o al efecto placebo, o dicho de otro modo, podemos preguntarnos ¿existe una relación causal entre el fármaco y la mejoría, o simplemente se debe al hecho de tratar? Surgirá, pues, la pregunta: ¿si se hubiese administrado otro tratamiento –o placebo–, qué hubiera ocurrido? Este modelo de pregunta es denominada en filosofía de la ciencia *condicional subjuntivo*<sup>18</sup>. En los ensayos clínicos usuales, para contestar a esta pregunta se hace necesario disponer de un grupo de

TABLA II  
REQUISITOS DE LA ENFERMEDAD Y TRATAMIENTOS

1. Que la enfermedad sea razonablemente estable en el tiempo, de modo que los posibles cambios bruscos en la enfermedad no confundán la evaluación	En este sentido las enfermedades agudas o rápidamente evolutivas no son susceptibles de ser estudiadas con EC n = 1
2. Que la acción de los tratamientos se alcance rápidamente	Fármacos o tratamientos que tarden en alcanzar efecto no son muy adecuados para estos estudios
3. Que al cesar el/los tratamientos la situación pueda considerarse idéntica al estado anterior, de modo que el segundo tratamiento actúe como si el primero no se hubiera administrado	Fármacos con largo período de eliminación (amiodarona) plantean problemas para realizar el EC n = 1. Procedimientos irreversibles (cirugía, angioplastia, etc.) no son adecuados para estos estudios
4. Que la variable respuesta sea de evaluación fácil y de pronta aparición	Para objetivos a medio o largo plazo no es pertinente (o posible) realizar estos estudios
5. Que el tratamiento deba mantenerse durante largo tiempo	En caso contrario, no es eficiente realizar el esfuerzo de un EC n = 1

EC n = 1: ensayos clínicos para un solo paciente.

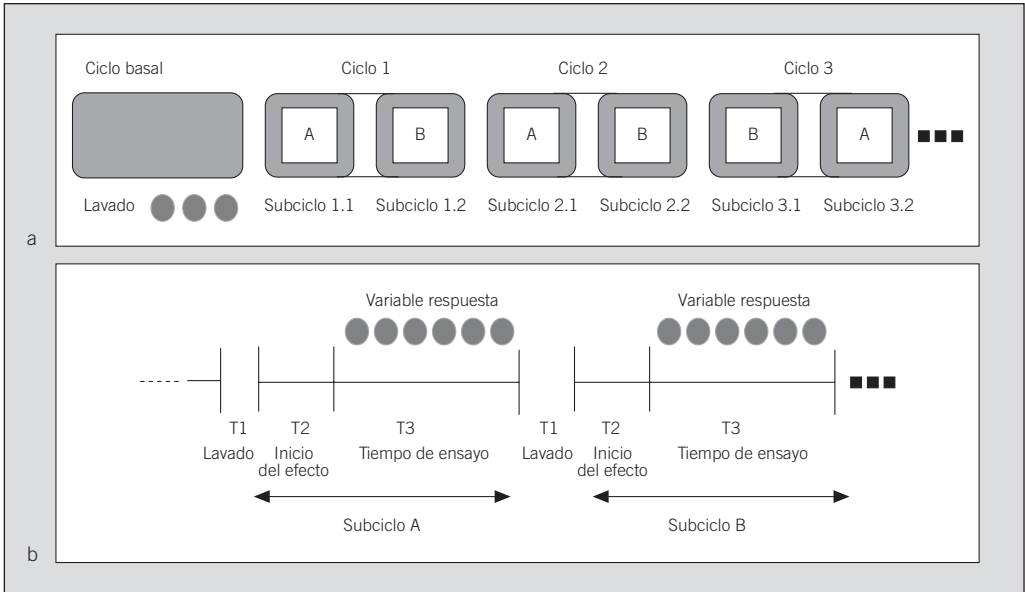


Fig. 1. a) Diseño de un EC  $n = 1$ . A: tratamiento A; B: tratamiento B. b) Detalle de un ciclo. T1: tiempo de lavado; T2: tiempo para el inicio del efecto; T3: tiempo de ensayo.

control tan parecido como sea posible al grupo de tratamiento, lo cual se consigue mediante la asignación aleatoria.

En nuestro caso, al tratarse de un solo paciente se plantea un problema con la selección del control; no obstante, si la enfermedad es crónica y estable, el tratamiento actúa con rapidez y al suspenderlo deja de actuar también rápidamente, se podrá comparar un paciente consigo mismo en dos momentos diferentes y con diferentes tratamientos. Sin embargo, podría ocurrir que influencias externas o ligeros cambios transitorios en la enfermedad afectaran a la comparación y produjeran que un tratamiento aparezca como mejor que otro por azar. Por ello, es necesaria la repetición de los ciclos de modo que la ventaja de un tratamiento sobre el otro no pueda ser atribuida al azar o a la casualidad.

Los requisitos de la enfermedad y de los tratamientos necesarios para realizar estos estudios se muestran en la tabla II; tales condiciones son comunes a los estudios autocontrolados y cruzados.

#### *Pregunta principal, variable/s respuesta*

Existirá un objetivo del estudio que será terapéutico, evaluable a corto plazo y no relacionado con el pronóstico a largo plazo (p. ej., mortalidad

coronaria). La pregunta principal se formulará de modo concreto y con contenido empírico definido. La formulación podrá hacerse como hipótesis de una o de dos colas. La variable resultado deberá ser explícita y evaluable. Por ejemplo, "en nuestro paciente, A es mejor/diferente que B para el tratamiento de su limitación crónica al flujo aéreo. El efecto lo mediremos mediante el cuestionario de Borg<sup>19</sup> para la disnea, el consumo de betaestimulantes y la calidad del sueño, medido con cuestionarios específicos al paciente y al familiar".

La variable de respuesta puede ser un valor analítico (p. ej., potasio o hemoglobina A<sub>1c</sub>), una imagen ecográfica o radiológica, un registro eléctrico, una prueba funcional, etc. No obstante, hay que señalar que estos ensayos se adaptan de modo especial a la valoración de síntomas y estados funcionales. Los posibles aspectos susceptibles de ser variable de respuesta de estos ensayos son innumerables y sólo limitados por la importancia que el paciente otorga al síntoma o problema concreto y la perspicacia del investigador clínico para la medición del síntoma o para la identificación y medición del constructo adecuado.

Desde un punto de vista teórico, la variable resultado de los EC  $n = 1$  podrá ser: a) *variable úni-*

ca: número de crisis epilépticas; b) variables múltiples: disnea, uso de betaestimulantes y alteraciones del sueño; en estos casos podrán ser analizadas de modo separado, jerarquizando la variable principal y las secundarias, y c) un constructo con múltiples componentes (actividades de la vida diaria, calidad de vida y cuestionario general de síntomas) o un índice clinimétrico construido con diversas variables. Esta tercera alternativa tiene interés porque puede ofrecer una perspectiva global del beneficio general/perjuicio de uno u otro tratamiento.

Es recomendable que la/s variable/s resultado sean medidas varias veces en cada subciclo para permitir una mejor evaluación del paciente y para mejorar el análisis estadístico del estudio.

### Ciclos y subciclos

En el EC  $n = 1$  se denomina ciclo al período de tiempo en el que se administran dos tratamientos a prueba, una vez cada uno, y se observan las variables de respuesta. Un ciclo está formado por dos subciclos, cada uno de los cuales corresponde a uno de los dos fármacos. La duración de cada subciclo de tratamiento y el número de ciclos a realizar constituyen aspectos prácticos "cruciales" en el diseño de los EC  $n = 1$ ; ambas cuestiones desempeñan un papel análogo al tamaño del estudio en los ensayos habituales, y se puede denominar *duración del estudio*.

*Período basal, o ciclo basal.* Siempre que sea clínicamente posible, es deseable la definición de un período de línea basal durante el que se efectúa el lavado de fármacos previos, se familiariza al paciente con el instrumento de medición, y se realizan las primeras mediciones repetidas. Es posible mantener medicaciones básicas durante todo el estudio y realizar el lavado sólo del tratamiento problema.

La definición de este período permite, así mismo, estabilizar las variaciones en la medición de la variable resultado y mejorar las posibilidades de análisis del estudio. Algunos autores<sup>10</sup> proponen dos subciclos de tratamiento abierto para identificar posibles efectos adversos y de este modo evitar las posibles retiradas de pacientes durante el estudio.

*Duración de los subciclos.* El tiempo de duración de cada subciclo se calculará sumando los tiempos de lavado (T1), el tiempo de comienzo del efecto (T2) y el tiempo de ensayo (T3).

a) T1. Para evitar el efecto de sobrearrastre es necesario programar entre cada subciclo una

fase de lavado de los fármacos en estudio. Para su cálculo, se usará la información farmacocinética disponible sobre el tiempo que tardan los fármacos en eliminarse y cesar su efecto; en caso de amplia variabilidad farmacocinética se usarán los tiempos máximos, y cuando los diferentes fármacos tengan diferente período de lavado consideraremos el mayor de ellos. Si es necesario se puede realizar previamente, en línea basal, una prueba de eliminación.

La posible contraindicación de esta fase, o del lavado del período basal, debe ser evaluada por el clínico en cada caso.

b) T2. Para el cálculo del mismo se usará, igualmente, el conocimiento de la farmacocinética. También aquí puede ser de utilidad la determinación previa de concentraciones del fármaco en el paciente.

c) T3. Durante el tiempo de ensayo se observarán los efectos y se medirán las variables respuesta. Este período se calculará de diferente modo según el tipo de variable.

Si la variable de respuesta es un índice de estado, o prueba analítica, la duración del componente T3 del subciclo se calculará intuitivamente, de modo que permita la realización de múltiples mediciones. Si la variable respuesta es un episodio ocasional, o periódico, la cuestión es más compleja: si la variable principal  $X =$  "número de episodios que ocurren en un intervalo temporal de duración  $t$ " se distribuye según un modelo de Poisson (que asume que los eventos son raros, aleatorios en el tiempo y con media constante), la variable  $T =$  "tiempo necesario hasta que ocurra el primer episodio" sigue un modelo exponencial<sup>21</sup>.

La función de distribución acumulativa del modelo exponencial es  $F(t) = 1 - e^{-\lambda t}$ , siendo  $\lambda$  el número medio de episodios que ocurren en la unidad de tiempo; por tanto, para calcular el tiempo para el cual hay una probabilidad prefijada " $p$ " de observar un episodio hay que resolver la ecuación  $p = 1 - e^{-\lambda t}$ , que puede expresarse como  $\lambda t = -\ln(1-p)$ , que para  $p = 0,95$  es aproximadamente  $t = 3/\lambda$ . Dicho de otro modo, si en promedio ocurre un episodio cada tiempo  $t$ , el tiempo necesario para tener una probabilidad de 0,95 de observar un episodio es  $3t$ .

Por tanto, desde el punto de vista práctico se tratará de ver si el episodio tiene cierta regularidad, y a continuación calcularemos con qué periodicidad se produce éste; si multiplicamos por tres ese período tendremos una seguridad del 95% de observar al menos un episodio durante el mismo (regla del 3). Ese tiempo será el componente T3 del subciclo en el EC  $n = 1$ . Si la fre-

cuencia media de aparición del episodio no es constante, no se dan las presunciones para el uso del modelo de Poisson, y además no se dan los requisitos de regularidad en la enfermedad o en su manifestación para plantear un  $EC\ n = 1$ .

**Número de ciclos.** El número de ciclos a programar en el ensayo dependerá del modo de evaluar el efecto: si se va a evaluar cada ciclo en términos de éxito/fracaso de un tratamiento, asumiendo independencia entre ciclos se podrá formular una diferencia de éxitos mínima para considerar un fármaco mejor que otro y estimar, a partir de la distribución binomial, el número de repeticiones necesarias. Esta estimación es, sin embargo, la más desfavorable desde el punto de vista de la duración del estudio, que para el mejor de los casos deberá durar 5 ciclos (véase "Análisis de datos").

Si la variable de respuesta es un constructo, índice clínico, o un valor analítico, y puede ser medida muchas veces en cada subciclo, el número de ciclos desempeña un papel menos crítico y se calculará intuitivamente.

En ambos casos es aconsejable fijar *a priori* el número de ciclos, aunque, si la respuesta es evidente, se pueda decidir el final del estudio. En este sentido, la experiencia publicada<sup>20</sup> indica que la realización de menos de 3 ciclos difícilmente aporta evidencias clínicas.

**Secuencia de tratamientos.** El orden de los tratamientos en cada ciclo se asignará de modo aleatorio. Existen varios procedimientos. El más fácil es lanzar una moneda para cada ciclo y el resultado decide el orden de los dos tratamientos; alternativamente se puede elegir una secuencia de números aleatorios y fijar según ella (p. ej., par/impar) el orden de todos los ciclos. Una opción interesante es generar una serie de permutaciones de dos elementos, ordenarlas y asignar a cada una un número aleatorio, y la secuencia se ordenaría mediante esos números; esta opción permite generalizar el procedimiento para más de dos tratamientos. En todo caso, para preservar el enmascaramiento del médico el proceso de aleatorización debe ser ejecutado y conservado por alguien ajeno al estudio (p. ej., unidad de investigación).

#### *Definición de la intervención y conducción del estudio*

La intervención debe definirse de modo estricto y explícito (dosis, vía, preparados, etc.), y los placebos o tratamientos de control cuidadosa-

mente preparados; en este sentido, la colaboración de los farmacéuticos del centro es imprescindible. Es posible realizar el ensayo manteniendo en todos los ciclos un tratamiento base orientado a otros objetivos y poner a prueba otros tratamientos, a condición de que no existan interacciones entre tratamiento base y tratamientos en prueba, por ejemplo, se puede mantener tratamiento con anticoagulantes mientras se prueba un fármaco antiarrítmico.

Cuando la evaluación se refiere a síntomas, el enmascaramiento es imprescindible tanto en el clínico como en el paciente: en el clínico, para garantizar la "objetividad" en la evaluación de las variables y para evitar que existan mejores cuidados paralelos en algunos períodos. Ambos son problemas potenciales derivados de preferencias del clínico (conscientes o no) por alguna hipótesis; en el paciente, para evitar el efecto placebo o preferencias *a priori* por alguno de los tratamientos, así como el sesgo de obsequiosidad o tendencia a mostrar comportamientos que agraden al médico. Cuando se trate de variables de más fácil medición (valor analítico o ecográfico) el enmascaramiento es menos relevante, aunque deseable.

Los criterios de conclusión del estudio y las razones de interrupción del ciego deben ser explícitos. Un apartado relevante de la conducción del estudio es la comprobación de la observancia del tratamiento y se realizará mediante las estrategias clásicas de recuento de pastillas y el análisis de las concentraciones plasmáticas del fármaco<sup>22</sup>; también es importante el mantenimiento de una relación "estrecha" de animación con el paciente. A este respecto, señalaremos que el  $EC\ n = 1$  es un pacto explícito entre médico y paciente en beneficio de este último.

#### *Análisis de datos*

El modo de analizar los resultados es, sin duda, el aspecto más complejo y polémico de estos ensayos. Las propuestas van desde el uso de intuitivas soluciones gráficas a procedimientos estadísticos con diferente nivel de complejidad. En este artículo nos limitaremos a plantear el problema del análisis y enunciar algunas soluciones.

Como en cualquier experimento, el análisis de resultados trata de determinar si existen cambios en las variables respuesta en las diferentes condiciones experimentales, o dicho con mayor precisión, la pregunta es si los cambios entre las condiciones son mayores que dentro de cada condición experimental, y en su caso si el resul-

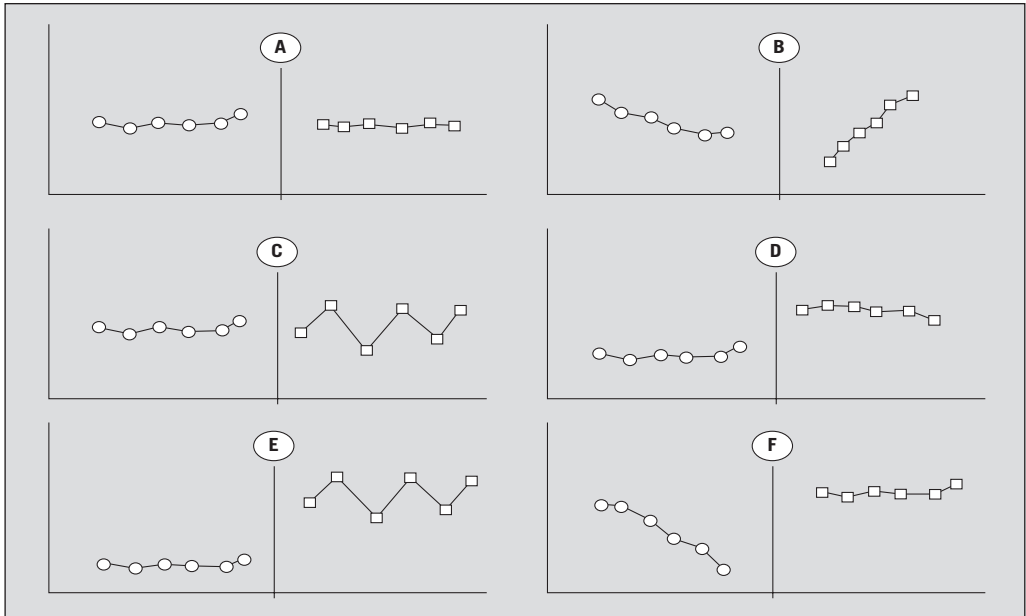


Fig. 2. Esquemas y criterios para el análisis gráfico. A: sin cambio entre 2 subciclos; B: cambio de tendencia; C: cambio de variabilidad; D: cambio de nivel; E: cambio de nivel y variabilidad; F: cambio de nivel y tendencia.

tado es repetible (de aquí la necesidad de varios ciclos).

Este enunciado general del análisis puede, no obstante, verse desde dos perspectivas: si consideramos el EC  $n = 1$  como un procedimiento para decidir el mejor tratamiento para un paciente concreto, el resultado del ensayo debe ser claro y comprensible tanto para el médico como para el paciente. Este es el argumento que anima las propuestas de expresar gráficamente los resultados y realizar un análisis visual<sup>23,24</sup>. El método consiste en presentar en un sistema de ejes cartesianos la variable resultado en función del tiempo, señalando en abscisas los diferentes ciclos y subciclos (fig. 2). Los criterios a considerar para el análisis visual son: a) la variabilidad en la medición general y entre ciclos, y b) cambios entre ciclos en la tendencia, en el nivel o en ambos<sup>24</sup>.

Además de la sencillez, la ventaja del método es la "insensibilidad" para efectos nimios<sup>25</sup>. Los inconvenientes residen en que la magnitud del efecto en una gráfica depende de la escala usada en las ordenadas, y que la interpretación presenta poca consistencia entre observadores<sup>23</sup> incluso con entrenamiento adecuado.

Como cuantificación de los cambios en cada ciclo se usa el denominado *split middle method*

(o línea de aceleración)<sup>26</sup>, consistente en dibujar una recta que pase por los puntos que representan la mediana del resultado en las dos mitades del ciclo basal; a continuación se cuenta el número de puntos por encima y por debajo de esa línea en cada uno de los subciclos de tratamiento. Si no existiera efecto del tratamiento, el número de puntos por encima y por debajo de la recta debería ser idéntico en ambos grupos de subciclos de tratamiento (subciclos A y subciclos B). Esta hipótesis se puede contrastar usando la distribución binomial y asumiendo que, en caso de existir tendencia, ésta es lineal; no obstante, la validez de esta prueba ha sido cuestionada<sup>27</sup>. Otro método gráfico análogo es la *resistant trend line* (o línea de resistencia de la tendencia)<sup>26</sup>, que usa tres puntos al calcular la línea de tendencia; este método es más poderoso que el anterior, pero necesita un mínimo de 12 determinaciones basales.

Si, por el contrario, abordamos el análisis desde perspectivas numéricas, los resultados no serán tan subjetivos; sin embargo, existirán dificultades de realización e interpretación de las pruebas estadísticas. Los procedimientos convencionales como el ANOVA o la prueba de la *t* de Student pueden ser inadecuados porque exi-

gen independencia entre las observaciones; en los EC  $n = 1$ , donde se realizan múltiples determinaciones seriadas de un mismo sujeto, esta condición, en teoría, no se cumple, pues las observaciones estarán correlacionadas entre sí (autocorrelación serial). La importancia de esta autocorrelación es también objeto de debate; algunos autores insisten en ella<sup>23,24</sup>, mientras que otros rechazan su existencia en la práctica<sup>11</sup>. Por otra parte, estas pruebas informan sobre los cambios en el valor medio de las variables, cuando en estos estudios puede ser más relevante el cambio de tendencia<sup>24</sup>.

Existen soluciones matemáticas para eludir esos problemas, la más característica de las cuales es el análisis de series temporales; este método cumple los supuestos estadísticos, pero su interpretación es compleja y precisa de un número de observaciones superior al usual en estos ensayos<sup>24</sup>. Otra alternativa propuesta es realizar un contraste de hipótesis para encontrar dependencia serial, y en caso de no encontrarla en los datos proceder con pruebas convencionales. Esta última propuesta es problemática puesto que usa a su favor la asimetría de los contrastes de hipótesis en situación de bajo tamaño muestral (por tanto, de baja potencia). De este modo, es probable no poder demostrar dependencia serial en las observaciones aun cuando exista.

Cuando la variable de respuesta es binomial, es decir éxito/fracaso del tratamiento A, la prueba de los signos puede ser adecuada. Con ella se calcula la probabilidad de la serie de respuestas obtenidas en la hipótesis de que la respuesta se deba al azar. En la hipótesis de que los dos tratamientos son iguales, la probabilidad de que un paciente prefiera el tratamiento A en 4 ciclos seguidos será  $1/2 \times 1/2 \times 1/2 \times 1/2 = 0,0625$ . El principal inconveniente es el número de ciclos necesarios para llegar al nivel de significación convencional, además de que esta prueba también exige independencia.

Una propuesta reciente<sup>28</sup>, particularmente útil si se usan distintas variables independientes, es definir un "índice beneficio/perjuicio" que englobe todas las variables e integre las dependencias temporales, y sobre ese índice aplicar las pruebas estadísticas o gráficas convencionales.

Hay que hacer notar, sin embargo, que la aparente solidez del análisis estadístico presenta dos puntos críticos: el primero se refiere a la pertinencia de usar números procedentes de escalas ordinales o casi dimensionales como si fueran proyecciones sobre la recta de los números reales, y el segundo se refiere a la escasa consistencia de los resultados según se usen unos u otros

métodos estadísticos, y que ha sido demostrada en los estudios de Nourbakhsh y Ottenbacher<sup>29</sup>.

En resumen, respecto al análisis de los EC  $n = 1$  existen problemas relacionados con la opinión sobre el objetivo del ensayo, y también dificultades de índole metodológica estadística. En todo caso, se ha producido una importante progresión en estos aspectos y ya existe un interesante catálogo de recursos a usar por el investigador. Por ello, la recomendación general es evitar un análisis estándar, diversificando el análisis mediante la combinación de métodos estadísticos con métodos gráficos, y adaptar el mismo a las peculiaridades de cada situación y a la comprensión del clínico y del paciente.

### Utilidad de los estudios

En este punto, es preciso plantearse: ¿en qué medida estos estudios son útiles para decidir sobre el tratamiento en pacientes, o en general sobre su manejo clínico?, y también cabe preguntarse ¿cómo pueden contribuir estrategias de este tipo a la mejoría de la evaluación de la efectividad de tratamientos en estudio?

Respecto a la primera cuestión, es obvio que no se trata de un diseño adecuado para todo tipo de pacientes, sino sólo una solución pertinente para pacientes determinados y que además reúnan algunos requisitos respecto a la enfermedad y a los tratamientos a aplicar. La experiencia del grupo de McMaster<sup>20,30,31</sup>, sobre 73 ensayos  $n = 1$ , indica que existe un amplio catálogo de enfermedades crónicas en las que han sido usados, aunque confirma que existen muchas situaciones en las que no es posible el uso. Así mismo, señala que existe un número de pacientes y clínicos (7 y 1, respectivamente) que rompen el protocolo del estudio, y en 8 pacientes el ensayo no se pudo comenzar o hubo de suspenderse precozmente por enfermedades o episodios intercurrentes.

Sin embargo, de los 57 ensayos concluidos en el 88 % se obtiene información clínica definitiva, aunque sólo en el 33 % del total se obtuvo significación estadística (el 35 % de los que ofrecen información definitiva), datos que refuerzan los comentarios realizados en el apartado del análisis en relación con el relativo valor de la estadística. Dentro de los ensayos concluidos, en el 37 % el resultado del EC  $n = 1$  indujo el cambio del tratamiento previsto inicialmente por el clínico, y en el 43 % el resultado indujo el mantenimiento del mismo. Tras el EC  $n = 1$  la confianza en la idoneidad del tratamiento fue alta o muy alta en la mayoría de los clínicos.



Otros estudios<sup>32,33</sup> indican resultados análogos respecto a la utilidad, y calculan los costes por ensayo<sup>32</sup> en 400 dólares de 1991, si bien no ofrecen un cálculo de los costes derivados de la interrupción de las medicaciones, de los gastos de monitorización de tratamientos y del tratamiento de los efectos adversos.

En resumen, cabe concluir que la realización de EC  $n = 1$  es un área de frontera que concilia la investigación y la acción en la clínica, que puede mejorar las acciones clínicas, particularmente en casos complejos, y que su realización pudiera ser coste-efectiva. Si el EC  $n = 1$  es una estrategia alternativa a la “prueba de tratamientos” habitual de la actividad clínica, ha de responderse mediante la comparación de ambas estrategias con diseños adecuados como los propuestos por Mahon et al<sup>34</sup>.

Respecto a la segunda pregunta, los EC  $n = 1$  incluidos en ensayos paralelos tradicionales pueden mejorar la evaluación de la efectividad del tratamiento en cada individuo de la población de estudio. En este sentido, una estrategia de múltiple cruzamiento y evaluación individual<sup>35</sup> ha sido utilizada en ensayos clásicos sobre insuficiencia cardíaca y digital en 1984. Recientemente se ha propuesto<sup>28</sup> el uso de un diseño de múltiples EC  $n = 1$  y un tratamiento estadístico especial para mejorar la evaluación de la efectividad en ensayos clínicos paralelos orientados hacia las variables de estado (calidad de vida y síntomas). Por tanto, los EC  $n = 1$  pueden también ser un instrumento para la progresión en la investigación y evaluación de fármacos o acciones terapéuticas.

### Aspectos bioéticos de los EC $n = 1$

Los ensayos clínicos implican la utilización de personas en la construcción de conocimiento y ello plantea un problema desde la visión kantiana de la dignidad del hombre como un fin en sí mismo<sup>36</sup>. Esta situación, inevitable en la investigación en humanos, plantea cuestiones bioéticas de interés.

El eje de la reflexión bioética asienta en los tres principios señalados por el informe Belmont<sup>37</sup>: *justicia*, *autonomía* y *beneficencia*. En algunas propuestas posteriores el principio de beneficencia puede contemplarse como subdividido en dos principios con diferente nivel de obligación<sup>38</sup>: serían los principios de “no maleficencia” y “beneficencia”. La jerarquía de principios podría plantearse en dos niveles de exigencia, un nivel superior donde se ubican los principios de “justicia” y “no maleficencia” (ambos exigibles), y un nivel inferior donde se ubicarían autonomía y beneficencia, principios jerarquizados por ese orden.

En la clínica, los actores de esos principios están claramente definidos; la autonomía sería el principio propio del paciente y la beneficencia el del médico, por su parte sociedad y administración velarían por los principios de “justicia” y de “no maleficencia”<sup>38</sup>. En el ensayo clínico, el principio de beneficencia se difumina, puesto que el posible beneficio para el paciente o para futuros pacientes sólo es una posibilidad teórica<sup>39</sup> (se trata de un tratamiento en prueba) y, por otra parte, existen posibilidades de perjudicarlo. Por ello, en un primer nivel los principios de justicia y no maleficencia deben ser preservados, pero el principio prevalente es el de autonomía, cuya expresión es el *consentimiento informado*. En España, los comités éticos de investigación tienen el encargo legal<sup>40</sup> de velar por el cumplimiento de estos principios.

En el EC  $n = 1$ , los beneficios obtenidos de la realización de un ensayo clínico son directamente aplicables a la persona objeto del estudio; por tanto, el principio de beneficencia recupera concreción y se mantiene vigente el principio de autonomía. Como se señaló, el EC  $n = 1$  sólo es posible concebido como un pacto explícito entre médico y paciente para conocer, con el mejor método posible, el mejor tratamiento para éste.

En nuestro país no existe derecho positivo en relación con los EC  $n = 1$ , excepción hecha de un comentario sobre el uso compasivo de fármacos en un solo paciente. En el mundo anglosajón existen pronunciamientos<sup>11</sup> de comités éticos de investigación que argumentan que cuando el medicamento ya está probado no es preciso el pronunciamiento explícito de ellos sino de un comité de ética clínica, opinión que es compartida con algunos autores<sup>41</sup>.

Finalmente, señalaremos que los EC  $n = 1$  suponen un desplazamiento en el centro de gravedad en la investigación terapéutica. En efecto, los síntomas serán el objetivo principal en estos ensayos, frente a los objetivos clásicos de duración de vida, etc. Este desplazamiento es producto de la emergencia del principio de autonomía y plantea, como reto para la investigación clínica, la necesidad de progresar en el desarrollo de técnicas para medición de estados y síntomas (clinimetría)<sup>42</sup>.

La emergencia del principio de autonomía no es un objetivo en sí mismo, pero con ella la relación médico-enfermo tiende a equilibrarse en perjuicio del tradicional paternalismo médico; ese equilibrio es esencial para la consecución de una relación médico-enfermo que supere el puro paternalismo y el puro autonomismo, para fundarse en lo que Laín Entralgo<sup>43</sup> denomina *amistad médica*.

## Agradecimiento

A los Dres. Jaime Latour, José J. Mira, César Picó y Francisco Pozo por sus valiosos comentarios a versiones previas de este manuscrito.

Así mismo, estamos en deuda con Alicia Picó por su ayuda en la edición del manuscrito.

## BIBLIOGRAFÍA

- Durán JA, Figuerola J. Cumplimiento de la medicación: características, factores que lo determinan y recomendaciones para manejarlo. *Med Clin (Barc)* 1988; 90: 338-343.
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Decidir la mejor terapéutica. En: Sackett DL, Haynes RB, Guyatt GH, Tugwell P, editores. *Epidemiología clínica. Ciencia básica para la medicina clínica (2ª ed.)*. Madrid: Editorial Médica Panamericana S.A., 1994; 191-248.
- Taylor DC. The components of sickness: diseases, illnesses and predicaments. *Lancet* 1979; 2: 1.008.
- Patel VL, Evans DA, Groen GJ. Biomedical knowledge and clinical reasoning. En: Evans DA, Patel VL, editores. *Cognitive Science in Medicine. Biomedical modeling*. Cambridge, Massachussets: Massachussets Institute of Technology Press, 1989; 53-85.
- Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *Br Med J* 1995; 310: 1.122-1.126.
- Guyatt GH, Sackett DL, Cook DJ. How to use an article about therapy or prevention. *JAMA* 1993; 270: 2.598-2.601.
- Angell M. Patient's preferences in randomized clinical trials. *N Engl J Med* 1984; 310: 1.385-1.387.
- Álvarez-Dardet C, Bolumar F, Porta Serra M. La medición de frecuencia de enfermedad. *Med Clin (Barc)* 1987; 88: 287-291.
- Kramer MS, Lane DA. Causal propositions in clinical research and practice. *J Clin Epidemiol* 1992; 45: 639-649.
- Guyatt GH, Sackett D, Adachi J, Roberts R, Chong J, Rosebloom D et al. A clinician's guide for conducting randomized trial in individual patients. *Can Med Assoc J* 1988; 139: 497-503.
- Guyatt GH, Sackett D, Taylor W, Chong J, Roberts R, Pugsley S. Determining optimal therapy—randomized trials in individual patients. *N Engl J Med* 1986; 314: 889-892.
- Keller JL, Guyatt GH, Roberts RS, Adachi JD, Rosebloom D. An N of 1 service: applying the scientific method in clinical practice. *Scand J Gastroenterol* 1988; 23 (Supl 147): 22-29.
- Barlow DH, Hersen M. Single-case experimental designs. *Arch Gen Psychiatry* 1973; 29: 319-325.
- Barlow DH, Hersen M. Single-case experimental designs. *Strategies for studying behavior change (2ª ed.)*. Nueva York: Pergamon Press, 1984.
- Kratchwill TR, editor. *Single subject research. Strategies for evaluating change*. Orlando: Academic Press, 1978.
- Kazdin AE. *Single-case research designs. Methods for clinical and applied settings*. Nueva York: Oxford University Press, 1982.
- Spilker B. Single-patient clinical trial. En: Spilker B, editor. *Guide to clinical trials*. Nueva York: Raven Press, 1991; 277-282.
- Wartofsky MW. *Introducción a la filosofía de la ciencia*. Madrid: Alianza Editorial, 1987; 326-329.
- Borg G. Perceived exertion as an indicator of somatic stress. *Scand J Rehabil Med* 1970; 23: 92-98.
- Guyatt GH, Keller JL, Jaeschke R, Rosebloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness. Our three years experience. *Ann Intern Med* 1990; 112: 293-299.
- Abraira V, Pérez de Vargas A. *Bioestadística*. Madrid: Ed. Centro de Estudios Ramón Areces, 1996.
- Privitera MD, Treiman DM, Pledger GW, Sahlroot JT, Handforth A, Linde MS et al. Dezinamide for partial seizures: results of an n-of-1 design trial. *Neurology* 1994; 44: 1.453-1.458.
- Ottenbacher KJ. Reliability and accuracy of visual analyzing graphed data from single-subject designs. *Am J Occup Ther* 1986; 40: 464-469.
- Wolery M, Harris SR. Interpreting results of single-subject research designs. *Phys Ther* 1982; 62: 445-452.
- Baer DM. Perhaps it would be better not to know everything. *J Appl Behav Anal* 1977; 10: 167-172.
- Ottenbacher KJ. Analysis of data in idiographic research. *Am J Phys Med Rehab* 1992; 71: 202-208.
- Crosbie J. The inability of the binomial test to control type I error with single-subject data. *Behav Assess* 1987; 9: 141-150.
- Bagne CA, Lewis RF. Evaluating the effects of drugs on behavior and quality of life: an alternative strategy for clinical trials. *J Consult Clin Psychol* 1992; 60: 225-239.
- Nourbakhsh MR, Ottenbacher KJ. The statistical analysis of single-subject data: a comparative examination. *Phys Ther* 1994; 74: 768-776.
- Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse T. Clinical usefulness of n-of-1 randomized trials in patients with nonreversible chronic airflow limitation. *Am Rev Respir Dis* 1991; 144: 962-964.
- Jaeschke R, Adachi J, Guyatt GH, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 n-of-1 randomized controlled trials. *J Rheumatol* 1991; 18: 447-451.
- Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2 year period. *JAMA* 1993; 270: 2.708-2.712.
- March L, Irwing L, Schwartz J, Simpson J, Chock C, Brooks P. n-of-1 trials comparing a non-steroidal antiinflammatory drug with paracetamol in osteoarthritis. *Br Med J* 1994; 309: 1.041-1.046.
- Mahon JL, Laupacis A, Donner A, Wood T. A randomized trial of n-of-1 trials versus conventional therapy. *Clin Res* 1993; 4: 180A.

35. Lee DC, Johnson RA, Bingham JB, Leahy M, Dinsmore RE, Goroll AH et al. Heart failure in outpatients. A randomized trial of digoxin versus placebo. *N Engl J Med* 1982; 306: 699-705.
36. Rawls J. *A theory of justice*, Cambridge Mass. Belknap Press of Harvard University Press 1971 [traducción española]. Madrid: FCE, 1979.
37. Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. Washington, US Government: OPRR reports, 1979.
38. Gracia D. *Fundamentos de bioética*. Madrid: Eudema Universidad, 1989; 103.
39. Hellman S, Hellman DS. Of mice but no men. Problems of the randomized clinical trial. *N Engl J Med* 1991; 324: 1.585-1.589.
40. Real Decreto 561/1993 de 16 de abril por el que se establecen los requisitos para la realización de Ensayos Clínicos con medicamentos. BOE, 13 de mayo de 1993; 14.346-14.364.
41. Irwing L, Glasziou P, March L. Ethics of n-of-1 trials. *Lancet* 1995; 345: 469.
42. Feinstein AR. *Clinimetrics*. New Haven: Yale University Press, 1987.
43. Laín Entralgo P. *El médico y el enfermo*. Madrid: Alianza Editorial, 1987.

---

# El placebo en ensayos clínicos con medicamentos

---

Fernando García-Alonso<sup>a</sup>, Eliseo Guallar<sup>b</sup>,  
Olav M. Bakke<sup>c</sup> y Xavier Carné<sup>d</sup>

<sup>a</sup>Centro Nacional de Farmacobiología. Instituto de Salud Carlos III. Madrid. <sup>b</sup>Departamento de Epidemiología y Estadística. Escuela Nacional de Sanidad. Instituto de Salud Carlos III. Madrid. <sup>c</sup>OBAK Pharma Consulting. Barcelona.

<sup>d</sup>Servicio de Farmacología Clínica. Hospital de la Vall d'Hebron. Barcelona.

## En qué consiste el efecto placebo y por qué utilizarlo

---

La utilización del placebo como herramienta metodológica en ensayos clínicos apenas tiene 60 años. El *primum non nocere* hipocrático ha predominado de tal manera en la terapéutica, que sólo a partir de los años cincuenta se planteó de forma explícita que los medicamentos, aparte de no hacer daño, debían ser eficaces<sup>1</sup>. Hubo que esperar hasta la introducción de la enmienda Kefauver-Harris a la Food, Drug and Cosmetic Act en 1962 para que la demostración de eficacia fuera exigida como requisito obligatorio para la aprobación de un nuevo medicamento<sup>2</sup>. En la actualidad, el ensayo clínico aleatorizado, con enmascaramiento doble ciego y controlado con placebo es el método estándar para la evaluación de la eficacia de las intervenciones farmacológicas<sup>3-5</sup>.

El efecto placebo ha sido definido como el cambio terapéutico en el estado del paciente que está causalmente conectado con el conocimiento (o la conciencia) personal que posee de encontrarse en una determinada situación clínica<sup>6</sup>. Un placebo es, por tanto, un procedimiento médico que no posee efecto fisicoquímico específico sobre la situación de dicho paciente<sup>6</sup>. En investigación médica, el procedimiento habitualmente empleado consiste en suministrar al paciente una sustancia farmacológicamente inactiva en el acto terapéutico.

El uso del placebo en la investigación clínica y las críticas que este uso ha motivado<sup>7-10</sup> deben encuadrarse en el ámbito de la metodología del ensayo clínico, entendido como el paradigma del método científico aplicado a la terapéutica. El uso de placebos permite discernir qué parte del efecto observado de la intervención se debe a su

efectividad farmacodinámica y qué parte se debe al efecto placebo. De hecho, uno de los primeros elementos de autocrítica de los pioneros de la evaluación científica de la terapéutica (durante los años cincuenta y sesenta) fue, precisamente, el reconocimiento de la existencia del efecto placebo y la administración de placebos al grupo control de los ensayos clínicos como metodología estándar para evaluar el efecto farmacodinámico<sup>11</sup>. A su vez, la utilización de un grupo control que recibe únicamente la sustancia inactiva puede utilizarse en muy diferentes contextos, desde el de una enfermedad crónica y grave en el que se evalúa un presunto tratamiento curativo, hasta el de una breve experiencia destinada a estudiar el efecto sobre un síntoma leve o una variable de efecto intermedia (p. ej., un parámetro bioquímico o fisiológico), o el de una intervención preventiva cuya supuesta eficacia es necesario evaluar a largo plazo.

Sin embargo, la utilización de sustancias placebo en la investigación médica se realiza también en otros contextos<sup>4</sup>. Así, puede utilizarse un placebo al inicio del estudio para permitir al investigador observar la estabilización de la enfermedad o para acostumbrar al paciente a la rutina de la investigación. También se emplea el placebo en la fase de "lavado" entre dos períodos de tratamiento activo en los diseños cruzados y en los estudios de búsqueda de la dosis óptima de un medicamento. Finalmente, el uso de placebos facilita el enmascaramiento del ensayo clínico, lo que evita la introducción de sesgos en el seguimiento y evaluación de los grupos de comparación.

## Componentes del efecto placebo

---

Esquemáticamente se puede afirmar que la respuesta de un paciente consciente a una intervención farmacológica es, como mínimo, la suma de la acción de cuatro componentes sobre el curso

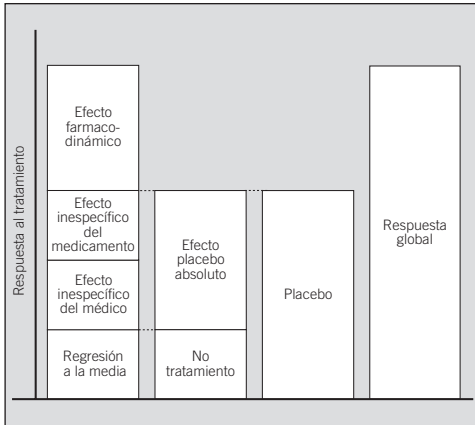


Fig. 1. Participación del placebo en la respuesta global al tratamiento. La respuesta global es la suma del efecto farmacodinámico más el efecto placebo, que a su vez es la suma del efecto placebo "absoluto" más el efecto de la regresión a la media. El efecto placebo "absoluto" es la suma del efecto inespecífico del medicamento y el del acto médico. La magnitud de cada uno de estos efectos en la figura es arbitraria y no representa la respuesta a ningún fármaco específico.

natural de la enfermedad (fig. 1): a) efecto farmacodinámico; b) efecto inespecífico debido al medicamento; c) efecto inespecífico debido al acto médico, y d) efecto de la regresión a la media.

A estos componentes cabría añadir un quinto elemento, el que se conoce como "efecto Hawthorne"<sup>12</sup>. Se trata de un fenómeno común a las ciencias sociales, descrito durante la década de los años treinta y que parte de la observación empírica de la mejoría que se deriva simplemente del hecho de que un individuo sea sometido a observación. Paradójicamente, el montaje que supone la objetivación de un fenómeno interfiere con el proceso que se quiere objetivar.

Es difícil conocer en qué proporción participa cada uno de estos elementos en una determinada observación, y puede ocurrir incluso que el efecto de alguno de ellos sea nulo. El efecto inespecífico debido al medicamento y el debido al acto médico constituyen lo que podría definirse como efecto placebo "absoluto" (fig. 1). A menudo es difícil distinguir entre regresión a la media y efecto placebo, por lo que, en la práctica, este fenómeno se considera un componente más del mismo<sup>13</sup>.

Mediante el diseño de un estudio de grupos paralelos con distribución aleatoria, enmascarado y comparado con placebo es posible disec-

cionar el efecto farmacodinámico del efecto placebo en la respuesta a una determinada intervención farmacológica. Más difícil resulta aislar cada uno de los tres componentes del efecto placebo descritos. La comparación del grupo tratado con placebo con un nuevo grupo control de no tratamiento permitiría distinguir el efecto placebo absoluto del efecto de la regresión a la media (fig. 1). Finalmente, para separar el efecto inespecífico debido al medicamento del debido al médico o a la interacción médico-paciente, bastaría introducir otro grupo control de no tratamiento, aunque esta vez con presencia activa del médico a lo largo de todo el procedimiento.

Aunque esta disección del placebo tiene más interés académico que práctico, puede ser útil en un ensayo clínico intentar reducir la variabilidad que introduce cada componente del efecto placebo. Por ejemplo, dado que el efecto inespecífico debido al medicamento puede variar en relación con la experiencia previa con medicamentos o procedimientos similares, cabría homogeneizar a los sujetos del estudio con respecto a su contacto previo con medicamentos o procedimientos de efecto terapéutico parecido. Para evitar diferencias en la respuesta por lo que respecta al efecto inespecífico atribuible al médico, bastaría con que éste fuera el mismo en todos los casos. Sin embargo, la variabilidad asociada a las experiencias del propio paciente sólo es controlable en el contexto de un diseño cruzado, en el que cada paciente recibe todas las opciones en estudio<sup>4</sup>. Finalmente, para minimizar el efecto de la regresión a la media, es aconsejable calibrar bien los aparatos de medida, disminuir la variabilidad inherente al proceso de medida y realizar varias determinaciones de las variables de medida antes de la aleatorización para obtener un valor basal estable<sup>13</sup>.

### Mecanismo de acción del efecto placebo

Se han descrito tres teorías en relación con los mecanismos mediante los que opera el efecto placebo: el modelo del condicionamiento y el modelo del significado<sup>5</sup>. La primera, más que una teoría, es un modelo que une la situación placebogénica con las propiedades analgésicas del efecto placebo y que parte de la observación de que la analgesia por placebo disminuye con la administración de naloxona. Más interés, para lo que aquí nos ocupa, tienen las otras dos propuestas. El modelo del condicionamiento se basa en la teoría de que diferentes personas (como médicos o enfermeras), lugares (hospitales o consultas), objetos (píldoras o in-

yecciones) o rituales, *a priori* neutrales, pueden condicionarse (como el perro de Pavlov) mediante la asociación con experiencias previas que, en la memoria del paciente, están ligadas a estas personas, lugares o rituales. El modelo del significado va más allá y argumenta que el efecto placebogénico proviene de los elementos culturales y simbólicos que rodean al acto terapéutico, sin necesidad de que exista previamente una experiencia condicionante.

Al evaluar un nuevo medicamento, interesa conocer su actividad terapéutica, independientemente del entorno y del médico que trate al paciente, lo sugestionable que éste sea y las oscilaciones propias de su enfermedad. Para ello, no existe otra fórmula que realizar, en algún momento del desarrollo clínico del fármaco, un estudio paralelo comparativo con un grupo sometido a placebo. Este tipo de diseño es el que está sujeto a una mayor polémica, dado que las agencias reguladoras lo consideran en muchas ocasiones preceptivo para demostrar la eficacia clínica, siempre y cuando no existan conflictos éticos relevantes que lo invaliden.

### **De la infrutilización al abuso del placebo**

La infrutilización del placebo en el desarrollo de medicamentos ha contribuido al actual desconocimiento del verdadero valor de una parte de nuestro arsenal terapéutico. Esta infrutilización se ha ido corrigiendo progresivamente, gracias, sobre todo, al impulso de la Food and Drug Administration y otras agencias de evaluación de medicamentos. En el fondo de la discusión subyace un problema de filosofía a la hora de autorizar nuevos medicamentos. Las reglas del libre mercado establecen que, para que un medicamento sea aprobado, debe demostrar ser eficaz y seguro en términos absolutos, y no necesita, por tanto, demostrar que es más eficaz y seguro que los ya existentes en aquel momento en el mercado. La eficacia comparativa en relación con su coste se deja para un segundo nivel de discusión, cuando se decida si ese nuevo medicamento será financiado o no por una determinada aseguradora sanitaria. Por tanto, las compañías farmacéuticas están interesadas en realizar en primer lugar estudios frente a placebo, para asegurar así su autorización y, después, frente a otros fármacos de referencia, para conseguir su selección frente a la competencia o, simplemente, su financiación.

Como el desarrollo de nuevos medicamentos se basa fundamentalmente en el hallazgo de pequeñas mejoras sobre lo ya establecido, la com-

paración frente a placebo permite hallar diferencias estadísticamente significativas con muestras de tamaño razonable, ya que amplifica la distancia entre un efecto y otro. Podría argumentarse que, idealmente, la comparación debería hacerse con un medicamento de referencia y una muestra suficientemente elevada, para, así, poder demostrar pequeñas diferencias<sup>7</sup>. Sin embargo, esta alternativa es tan costosa que muchas compañías farmacéuticas no podrían abordarla. Se daría el caso, además, de que miles de pacientes podrían ser sometidos a una terapia que, finalmente, no fuera mejor que el placebo. Precisamente este punto constituye la piedra angular de la discusión sobre el problema ético de la utilización de placebos.

Por otro lado, la utilización de un grupo placebo, además de un grupo control activo, es una excelente herramienta para la validación interna de un ensayo clínico, especialmente cuando no se encuentran diferencias entre el tratamiento experimental y el control activo, y queremos discernir si nos encontramos ante un resultado realmente negativo o ante un problema de diseño inadecuado.

Aunque es indudable, desde el punto de vista histórico, la aportación del placebo para demostrar la eficacia de un gran número de intervenciones terapéuticas, una inadecuada utilización del mismo en algunos casos recientes ha desatado toda una serie de críticas<sup>7-10</sup>. Los argumentos en contra del abuso del placebo se centran en el sometimiento de pacientes a una terapia con placebo cuando existen alternativas terapéuticas validadas. No estamos hablando de situaciones extremas, como los casos, ya históricos, de la penicilina frente a placebo en las infecciones respiratorias causadas por estreptococos (1996)<sup>1</sup>, los anticonceptivos orales frente a placebo en la prevención del embarazo (1971)<sup>1</sup>, o la ivermectina frente a placebo en la oncocercosis (1985)<sup>2</sup>, sino de situaciones hoy día muy frecuentes en las que se plantea el desarrollo de un nuevo medicamento, como son los casos de los nuevos antieméticos, antiulcerosos, antimigrañosos, antidepresivos, antihipertensivos o de los fármacos para tratar la insuficiencia cardíaca congestiva (ICC).

Es evidente que para todas estas enfermedades existen diversos tratamientos suficientemente validados y que, por tanto, el uso de placebos en el grupo de comparación en diseños paralelos debería quedar proscrito. Sin embargo, las cosas no son tan simples y cada caso merece un estudio detallado y una decisión específica. Algunos ejemplos pueden contribuir a aclarar la discusión.

## Antieméticos

Los ensayos clínicos de ondansetrón han sido el principal desencadenante de la reciente oleada de críticas contra el uso del placebo<sup>7-10</sup>; en un estudio realizado en pacientes con vómitos inducidos por cisplatino<sup>14</sup> se fue demasiado lejos utilizando placebo como comparación, ya que el ondansetrón había demostrado previamente su eficacia frente a la metoclopramida en ensayos clínicos aleatorizados<sup>15</sup>.

Sin embargo, esta situación no puede compararse, a nuestro juicio, con la utilización de placebo en un ensayo de búsqueda de dosis con ondansetrón oral en monodosis para la profilaxis de los vómitos posquirúrgicos<sup>16</sup>, ya que en este caso no existía un acuerdo sobre la pauta de dosificación e incluso muchos cirujanos y anestesiólogos eran partidarios de no realizarla. Han sido cientos de miles los pacientes que han sufrido por esta incertidumbre terapéutica, lo que hace que, en este caso, la utilización del placebo sea más un requisito ético que lo contrario. No debemos olvidar que la historia de la farmacología clínica está repleta de pautas posológicas equivocadas por no haber utilizado en su momento los diseños adecuados para dilucidarlas<sup>4</sup>.

## Antidepresivos

Debido a que en la actualidad existen múltiples antidepresivos de eficacia clínica demostrada (de acuerdo con una visión simplificada del problema), no sería éticamente aceptable utilizar placebo en ensayos clínicos en la depresión. Sin embargo, hay que distinguir entre la depresión mayor (diagnosticada según criterios estandarizados), en la que existe un riesgo de suicidio y para la que se han descrito pautas terapéuticas bien establecidas, y toda una gama de cuadros depresivos de diferente intensidad sobre los que existen grandes incertidumbres terapéuticas.

En la práctica clínica las situaciones intermedias son, con mucho, las más frecuentes. En los estudios de utilización de antidepresivos se observa que, en atención primaria, sólo una tercera parte de los pacientes son tratados al menos durante 4 semanas y que más de dos terceras partes de ellos reciben dosis infraterapéuticas<sup>17,18</sup>. Esto puede ser el reflejo de una mala práctica, pero también de ciertas lagunas de conocimiento que existen sobre la enfermedad y su tratamiento. Por ejemplo, ante un paciente que padece la llamada depresión doble<sup>19</sup>, existe incertidumbre acerca de si debe realizarse o no un tratamiento de fondo y en caso de hacer-

lo, se desconoce cuál es el agente de elección y cuáles deben ser su duración y su posología óptimas.

Con este planteamiento más amplio, un comité ético de investigación clínica debería revisar cuidadosamente y, salvo circunstancias excepcionales, rechazar un ensayo clínico frente a placebo en pacientes con depresión grave y riesgo de suicidio. Por el contrario, la utilización de placebo se convierte en una necesidad si queremos desentrañar, por ejemplo, los secretos de la depresión doble. Es decir, depresión y utilización de placebo no sólo no son antagónicos, sino que en muchas ocasiones pueden ser complementarios.

## Medicamentos para la insuficiencia cardíaca congestiva

Actualmente, los inhibidores de la enzima convertidora de la angiotensina (IECA) son parte del tratamiento estándar de la ICC<sup>20</sup>, lo que no implica necesariamente que ya no puedan hacerse ensayos clínicos frente a placebo en esta enfermedad. Precisamente ha sido en esta área de la terapéutica donde el ensayo frente a placebo ha hecho una de sus mayores contribuciones: el uso indiscriminado de los digitálicos durante 200 años sólo pudo ser acotado mediante un estudio en el que no pudo diferenciarse el efecto de la digitoxina y el del placebo en pacientes con insuficiencia cardíaca y ritmo sinusal<sup>21</sup>. Posteriormente otros investigadores pudieron confirmar estos resultados en estudios aleatorizados, con lo que, a partir de los años setenta, sólo se considera aceptable tratar con digital a los pacientes con ICC que cumplan determinados criterios. Estos estudios frente a placebo pudieron realizarse pese a que, hasta entonces, ¡la digital era el tratamiento estándar de la insuficiencia cardíaca!

Pero, aún más, varios fármacos que inicialmente prometían eficacia en la ICC han sido descartados por su toxicidad gracias a los estudios frente a placebo. Es posible criticar<sup>7</sup> un estudio de flosequinan frente a placebo en la ICC<sup>22</sup> que comenzó cuando ya existían indicios de que los IECA eran el tratamiento de referencia, pero precisamente gracias a un estudio en el que se utilizó placebo pudo dilucidarse el efecto tóxico del flosequinan, un medicamento que aumenta la mortalidad de los pacientes con ICC. De hecho, un estudio previo del flosequinan frente al captopril en pacientes con ICC grave había llegado a la conclusión de que ambos fármacos tenían una eficacia similar<sup>23</sup>.

Aunque actualmente ya no es aceptable privar a un paciente con ICC de grados III-IV de la clasificación de la New York Heart Association del tratamiento con IECA, existen otras circunstancias en las que hacerlo puede estar plenamente justificado, por ejemplo, en los casos en los que se evalúa un nuevo fármaco para el que se pretende probar su eficacia en los casos en los que falla la terapéutica estándar.

### *Antiarrítmicos*

En el estudio Survival with Oral *d*-sotalol (SWORD)<sup>24</sup> se comparó la eficacia del *d*-sotalol, un antiarrítmico que actúa bloqueando los canales de potasio, frente a placebo en pacientes con disfunción ventricular asociada a infarto de miocardio. Este estudio fue interrumpido prematuramente debido a un aumento de la mortalidad del 65 % en el grupo tratado con *d*-sotalol. Incrementos similares de la mortalidad se han observado en ensayos clínicos con encainida<sup>25,26</sup>, flecainida<sup>25,26</sup> y moricizina<sup>27</sup>, realizados cuando alguno de estos fármacos ya estaba aprobado como antiarrítmico y su uso era habitual en clínica. Según una visión restrictiva del uso del placebo, no era ético comparar fármacos de eficacia antiarrítmica demostrada frente a placebo. Sin embargo, tan sólo el uso de un grupo control tratado con placebo ha permitido demostrar de forma inequívoca que estos fármacos aumentan la mortalidad, convenciendo a la comunidad científica de la necesidad de restringir el uso de antiarrítmicos. En este caso, el uso de placebo ha sido beneficioso tanto para los pacientes que participaron en los ensayos y fueron asignados al grupo control como para futuros pacientes susceptibles de recibir terapia antiarrítmica.

### **Placebo y consentimiento informado**

Si en un ensayo clínico frente a placebo es previsible que un grupo de pacientes vaya a sufrir un daño relevante como consecuencia de privarle de un tratamiento que se considera necesario, el hecho de que los participantes hayan dado previamente su consentimiento informado no hace válida éticamente la experiencia. En este caso, aunque se haya respetado el principio de autonomía, se ha conculcado uno previo y más importante, el principio de no maleficencia. En sentido contrario, aunque el diseño del estudio sea correcto y no sea previsible que el paciente sufra un daño relevante como consecuencia de privarle de un determinado tratamiento, sigue siendo

obligatorio solicitar al paciente su consentimiento informado y comunicarle la naturaleza de las intervenciones en la hoja de información que debe acompañar a éste. En definitiva, en todo ensayo frente a placebo son exigibles tanto la garantía razonable de que el paciente no sufrirá un daño como el consentimiento informado previo del participante.

Hasta aquí hemos visto varios ejemplos de la valoración de la utilización del placebo en la investigación clínica, y hemos podido comprobar que existen muchas circunstancias clínicas en las que, aunque exista una terapia establecida para la enfermedad en estudio, no es previsible que se produzca un daño relevante derivado de la participación en un estudio controlado con placebo. Mucho más complicado resulta hacer llegar al sujeto de la investigación la información acerca de la lógica del placebo en la experiencia concreta que se le propone<sup>28,29</sup> para que dé un consentimiento verdaderamente informado. En ocasiones, es el propio médico quien se siente confuso e inseguro sobre esta cuestión. Sin embargo, estas circunstancias no eximen de la obligación de intentar proporcionar al participante una información lo más ajustada posible que tenga como objetivo conseguir el máximo de comprensión por su parte.

### **El placebo en sus justos términos**

A pesar de que existen una serie de recomendaciones destinadas a evaluar si la utilización del placebo como medicación única es aceptable en investigación terapéutica, estas recomendaciones son excesivamente genéricas<sup>5</sup>. En la tabla I se propone un decálogo de criterios para el uso del placebo en investigación clínica que intenta compaginar el respeto a los principios éticos individuales con las ventajas prácticas que, en el desarrollo de nuevos fármacos, supone la comparación frente a placebo y, por ende, el beneficio que obtiene de ello la sociedad.

Con los argumentos presentados a lo largo de este trabajo, se ha pretendido demostrar que, incluso con el actual desarrollo de nuestro arsenal terapéutico, la utilización de placebos en numerosas circunstancias no vulnera los principios éticos de la investigación clínica y, además, es un requisito indispensable para determinar la eficacia de las intervenciones farmacológicas. Finalmente, la utilización prudente e informada del decálogo de criterios presentado puede facilitar el uso del placebo en sus justos términos, huyendo tanto de la infrutilización como del abuso.



TABLA I  
DECÁLOGO DE CRITERIOS PARA EL USO DEL PLACEBO EN INVESTIGACIÓN CLÍNICA

<ol style="list-style-type: none"> <li>1. El uso de placebo en investigación clínica debe realizarse en el contexto de ensayos clínicos formales, cumpliendo los requisitos legales que rigen el desarrollo de nuevas entidades químicas o biológicas en seres humanos</li> <li>2. El uso de placebo debe estar justificado explícitamente en el protocolo del estudio, que además debe incluir una descripción detallada de los posibles riesgos y beneficios asociados a su uso</li> <li>3. La hoja de información al paciente del consentimiento informado del estudio debe describir explícitamente al placebo como una sustancia inactiva empleada como comparación</li> <li>4. Los ensayos clínicos frente a placebo de tamaño muestral moderado o amplio, o de duración prolongada, deben incluir en el protocolo un plan de análisis intermedios, que deben ser revisados periódicamente por un comité de monitorización externo</li> <li>5. La administración de fármaco activo es obligatoria siempre que se haya demostrado de forma concluyente la eficacia del fármaco o que la no administración del mismo pueda resultar en daño o lesión permanente. Sin embargo, puede justificarse el uso de placebo cuando el tratamiento estándar se asocie a una frecuencia elevada de efectos adversos graves</li> </ol>	<ol style="list-style-type: none"> <li>6. La utilización de placebo en sustitución de fármacos eficaces durante períodos limitados de tiempo puede justificarse en estudios de búsqueda de dosis, durante los períodos de preinclusión, en las fases de lavado de los ensayos cruzados o en la evaluación a corto plazo de la eficacia de un fármaco sobre variables intermedias, siempre y cuando la no administración del fármaco activo no resulte previsiblemente en daño o lesión permanente</li> <li>7. La demostración de eficacia de un determinado fármaco sobre variables intermedias no impide necesariamente la utilización del placebo en la evaluación de variables de supervivencia</li> <li>8. En intervenciones preventivas, el uso de placebo será aceptable siempre que los fármacos activos no hayan demostrado su eficacia sobre variables de supervivencia u otras de gran relevancia clínica</li> <li>9. Es justificable, desde el punto de vista ético, la adición de un placebo al tratamiento estándar para la evaluación de nuevos tratamientos</li> <li>10. La utilización de placebo en ensayos clínicos con voluntarios sanos y en estudios farmacocinéticos es perfectamente aceptable</li> </ol>
--	--

## BIBLIOGRAFÍA

1. Lachaux B, Lemoine P. Placebo, un medicamento que busca la verdad. Madrid: Interamericana-McGraw-Hill, 1989.
2. Lobo F. Medicamentos, política y economía. Barcelona: Masson, 1992.
3. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials (3.<sup>a</sup> ed.). St. Louis: Mosby-Year Book, 1996.
4. Bakke OM, Carné X, García-Alonso F. Ensayos clínicos con medicamentos. Barcelona: Mosby/Doyma Libros, 1994.
5. Collier J. Confusion over use of placebos in clinical trials. *Br Med J* 1995; 31: 821-822.
6. Hróbjartsson A. The uncontrollable placebo effect. *Eur J Clin Pharmacol* 1996; 50: 345-348.
7. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994; 331: 394-398.
8. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled, trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *Br Med J* 1995; 311: 844-846.
9. Soriguer Escofet FJC. Uso y abuso del placebo. *Med Clin (Barc)* 1995; 104: 576-577.
10. Rothman KJ. Placebo mania. *Br Med J* 1996; 313: 3-4.
11. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials of assess equivalence: the importance of rigorous method. *Br Med J* 1996; 313: 36-39.
12. Parsons HM. What happened at Hawthorne? *Science* 1974; 183: 922-932.

13. Guallar E, Jiménez FJ, García-Alonso F, Bakke O. La regresión a la media en la investigación y práctica clínica. *Med Clin (Barc)* 1997; 109: 23-26.
14. Beck TM, Ciociola AA, Jones SE, Harvey WH, Tchekmedyan NS, Chang A. Efficacy of oral ondansetron in the prevention of emesis in outpatients receiving cyclophosphamide-based chemotherapy. *Ann Intern Med* 1993; 118: 407-413.
15. Citron ML. Placebos and principles: a trial of ondansetron. *Ann Intern Med* 1993; 118: 470-471.
16. Rust M, Cohen LA. Single oral dose ondansetron in the prevention of postoperative nausea and emesis. *Anaesthesia* 1994; 49 (Supl): 16-23.
17. Laporte JR, Figueras A. Placebo effects in psychiatry. *Lancet* 1994; 344: 1.206-1.209.
18. MacDonald TM, McMahon AD, Reid IC, Fenton GW, McDevitt DG. Antidepressant use in primary care: a record linkage study in Tayside, Scotland. *Br Med J* 1996; 313: 860-861.
19. Berkow R. *El Manual Merck de diagnóstico y terapéutica*. Barcelona: Mosby/Doyma Libros, 1994.
20. Braunwald E. ACE inhibitors—a cornerstone of the treatment of heart failure. *N Engl J Med* 1991; 325: 351-353.
21. Starr I, Luchi RJ. Blind study on the action of digi-toxin on elderly woman. *Am Heart J* 1969; 78: 740-751.
22. Packer M, Narahara KA, Elkayan U, Sullivan JM, Pearle DL, Massie BM et al. Double-blind, placebo-controlled study of the efficacy of flosequinan in patients with chronic heart failure. *J Am Coll Cardiol* 1993; 22: 65-72.
23. Cowley AJ, Wynne RD, Swami A, Birkhead J, Skene A, Hampton JR. A comparison of the effects of captopril and flosequinan in patients with severe heart failure. *Cardiovasc Drug Ther* 1992; 6: 465-470.
24. Waldo AL, Camm AJ, DeRuyter H, Friedman PL, MacNeil DJ, Pauls JF et al. Effect of *d*-sotalol on mortality in patients with left ventricular dysfunction after recent and remote myocardial infarction. *Lancet* 1996; 348: 7-12.
25. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; 321: 406-412.
26. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. *N Engl J Med* 1991; 324: 781-788.
27. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* 1992; 327: 227-233.
28. Gøtzsche PC. Is there logic in the placebo? *Lancet* 1994; 344: 925-926.
29. Guerra D, Porta M, Camí J. La relevancia clínica del efecto placebo. *Med Integral* 1986; 7: 17-23.

---

# Utilidad de la incorporación de análisis de evaluación económica en los protocolos de los ensayos clínicos

---

Javier Soto

Unidad de Farmacoeconomía. Productos Roche S.A. Madrid.

La reciente publicación en su Revista de un ensayo clínico controlado (ECC)<sup>1</sup> es una magnífica ocasión para valorar las aplicaciones y utilidades prácticas de la integración de evaluaciones económicas de medicamentos (EEM) (o estudios de farmacoeconomía) en el diseño y realización de los protocolos de los ECC.

Los resultados de los ECC aportan información de indudable valor, como es la eficacia, la seguridad y la calidad de los productos estudiados. Estos datos van a servir tanto para su registro y aprobación como para decidir qué medicamento debe emplearse para tratar cada patología acorde a su momento evolutivo y según los diferentes pacientes.

Sin embargo, las conclusiones que se extraen de los protocolos clásicos de ECC no dicen nada sobre un aspecto cada día más importante, como es la eficiencia de las diferentes opciones terapéuticas evaluadas, esto es, el consumo de recursos que ha sido necesario emplear para conseguir los resultados clínicos obtenidos con los distintos tratamientos estudiados. Por lo tanto, un fármaco será más eficiente cuanto mejores resultados clínicos proporcione, con el menor coste posible<sup>2</sup>.

La incorporación de análisis de EEM en los ECC va a permitir determinar qué alternativa de las estudiadas presenta una mejor relación coste-efectividad y, por lo tanto, cuál va a ser más eficiente. La mayor aplicación práctica de este tipo de estudios va a ser la de ayudar a tomar decisiones racionales, justas y equitativas, a los diferentes agentes involucrados en esta tarea. De esta manera, la asignación de los recursos disponibles para el capítulo de gasto farmacéutico se podrá realizar de una manera más lógica, primándose el uso de los medicamentos más eficaces, lo que contribuirá a conseguir una mayor eficiencia del sistema sanitario<sup>3</sup>.

La inclusión de EEM durante las fases I y II puede ayudar a evaluar la conveniencia de proseguir con el desarrollo de un producto (estimando su probable eficiencia frente a las alternativas ya existentes) y a planificar las líneas futuras de investigación, y además sirve para sentar las bases que ayuden a su correcta implantación durante las fases de desarrollo más tardías (diseño de escalas específicas de calidad de vida, realización de estudios de coste de la enfermedad a la que va destinado el nuevo fármaco, validación de instrumentos y cuestionarios para recoger el consumo y la utilización de recursos, entre otros).

Durante la fase III, su incorporación va a permitir disponer de datos de eficiencia en el momento de su comercialización, lo que puede ayudar en el proceso de fijación del precio y a la hora de decidir si va a ser reembolsable con cargo al erario público. Teniendo en cuenta las limitaciones de los ECC a la hora de poder generalizar sus resultados, y que tanto la eficacia encontrada como los recursos consumidos durante su realización pueden no ser representativos de lo que sucederá en la práctica asistencial diaria, será necesario valorar cuidadosamente los datos (económicos y de eficacia) que será necesario recoger en el transcurso de su realización, con el objetivo de intentar aproximarlos a lo que pasará en condiciones de práctica habitual, tras la comercialización del nuevo fármaco<sup>4,5</sup>.

Posteriormente, a lo largo de la fase IV (y en ocasiones ya durante la fase IIIb), la integración de los EEM dentro de los ECC naturalísticos (que intentan parecerse a lo que sucede en el mundo real de la asistencia médica) va a permitir conocer la relación coste-efectividad del nuevo medicamento frente a las opciones ya existentes (y que más se emplean) cuando se utilizan en condiciones de uso rutinario.

Así, los diferentes agentes decisores (médico prescriptor, comité de farmacia y terapéutica, far-

---

Publicado en *Med Clin (Barc)* 1998; 110 (4): 157-158.

macéuticos de atención primaria, gerentes, directores médicos, directores de gestión y pacientes) podrán conocer las ventajas económicas (y/o de calidad de vida) del nuevo medicamento frente al resto de opciones disponibles. Esta información será imprescindible para su inclusión en los formularios, protocolos y guías de práctica clínica, creados con el fin de racionalizar el uso de los medicamentos<sup>6</sup>. Teniendo en cuenta que no es viable la inclusión de los EEM en todos los ECC, para que su incorporación fuera recomendable, deberían cumplirse unos criterios<sup>7</sup>.

Volviendo al estudio objeto de esta reflexión, con los resultados obtenidos es difícil decidir cuál de los dos medicamentos debería emplearse habitualmente para tratar la gonartrosis activa. Si solamente se valoran los datos de eficacia y seguridad encontrados, probablemente daría igual emplear uno u otro. Sin embargo, si dispusiéramos del dato de la eficiencia de cada tratamiento, se podría tomar una decisión más racional.

En este caso concreto, y dado que la eficacia obtenida es similar con ambos medicamentos, debería haberse realizado un análisis de minimización de costes, recogiendo todos los costes directos (coste de adquisición de la medicación, días de hospitalización, visitas a servicio de urgencias, visitas a especialistas y/o generalistas, pruebas complementarias realizadas, tratamiento de las reacciones adversas aparecidas, viajes a los centros sanitarios y cuidados necesitados en el domicilio de los pacientes) y todos los costes indirectos (días de baja laboral) generados por el uso de ambos fármacos. De esta manera, se hubiera conocido el coste incremental derivado de emplear uno u otro medicamento, lo que habría reflejado la eficiencia de ambos tratamientos.

En otros casos, y dependiendo de los medicamentos en evaluación y estudio y la enfermedad para la que van a ser usados, será necesario emplear otro tipo de análisis: coste-efectividad, coste utilidad o coste-beneficio, pudiéndose incluso incorporar más de uno, lo que va a proporcionar información válida desde diferentes ópticas y perspectivas<sup>8</sup>.

Si se pretende que los resultados de los EEM sean útiles en la toma de decisiones, éstos deberán ser creíbles, válidos y estar disponibles en el momento oportuno<sup>9</sup>. Por lo tanto, es preciso que se emplee una correcta metodología en su diseño, elaboración y realización, conjuntamente con un estricto análisis e interpretación de sus resultados y una total transparencia y objetividad en su divulgación<sup>10</sup>. Para que esto fuera una realidad, debería propiciarse la puesta en marcha de

unas normas de buena práctica económica, lo que sería garantía de una correcta realización y aplicación.

En la actualidad, y teniendo en cuenta la constante y creciente limitación de recursos disponibles para el cuidado sanitario en nuestro país, es crucial poder disponer de datos de eficiencia de los diferentes medicamentos disponibles en el mercado, ya que esto permitirá emplear los que presenten una mejor relación coste-efectividad (y/o una mejor calidad de vida).

Los análisis de los EEM deben ser entendidos como un complemento o extensión de la investigación clínica, y nos ayudarán a realizar una asignación de los recursos disponibles de una manera más razonable y lógica, con lo que se facilitará, además, la puesta en marcha de nuevas actividades en los centros sanitarios. De esta manera, se podrá mantener (e incluso incrementar) la calidad asistencial de los pacientes, que a la postre debe ser lo más importante para cualquier sistema de salud.

## BIBLIOGRAFÍA

1. Gijón Baños J y Grupo de Estudio de Nabumetona en Artrosis de Rodilla. Eficacia y seguridad de la nabumetona en el tratamiento de la gonartrosis: ensayo clínico comparativo con aceclofenaco. *Med Clin (Barc)* 1997; 109: 130-134.
2. Sacristán JA, Soto J, Reviriego J, Galende I. Farmaeconomía: el cálculo de la eficiencia. *Med Clin (Barc)* 1994; 103: 143-149.
3. Rutten F. Economic evaluation and health care decision-making. *Health Policy* 1996; 36: 215-229.
4. Baltussen R, Ament A, Leidl R. Making cost assessment based on RCTs more useful to decision-makers. *Health Policy* 1996; 37: 163-183.
5. Ellwein LB, Drummond MF. Economic analysis alongside clinical trials. Bias in the assessment of economic outcomes. *Int J Tech Assess Health Care* 1996; 12: 691-697.
6. Powe NR, Griffiths RI. The clinical-economic trial: promise, problems and challenges. *Controlled Clin Trial* 1995; 16: 377-394.
7. Haycox A, Drummond M, Walley T. Pharmacoeconomics: integrating economic evaluation into clinical trials. *Br J Clin Pharmacol* 1997; 43: 559-562.
8. Freud DA, Dittus RS. Principles of pharmacoeconomic analysis of drug therapy. *Pharmacoeconomics* 1992; 1: 20-32.
9. Drummond M. Economic analysis alongside clinical trials: problems and potential. *J Rheumatol* 1995; 22: 1.403-1.407.
10. Chi Tsuen Lin M, Vacani P, Drummond MF. Economic analysis with clinical trials. *Applied Clin Trials* 1995; 5: 60-66.

---

# Eficiencia de los investigadores en la selección de los enfermos para ensayos clínicos: a propósito de un estudio multinacional

---

Rafael Dal-Ré, Rafael Ortega y Julián Espada

Departamento Médico. SmithKline Beecham Pharmaceuticals. Madrid.

El desarrollo clínico de nuevos fármacos es una labor realizada por la industria farmacéutica<sup>1</sup>. Es un proceso que, por término medio, dura aproximadamente 6 años desde que se inician los estudios en fase 1 hasta la conclusión de los ensayos en fase 3<sup>2</sup>. La mayor parte de las compañías no se plantean unos calendarios preestablecidos para realizar la investigación y desarrollo completo de un nuevo medicamento; en la fase clínica, por el contrario, sí se suelen establecer unos plazos de realización ajustados<sup>3</sup>. Desde hace años, las compañías farmacéuticas están realizando muchos esfuerzos para acortar el tiempo requerido para llevar a cabo la fase de desarrollo clínico de nuevos compuestos<sup>4</sup>. En este sentido, el desarrollo clínico puede ser considerado como un proceso en el que es posible diseñar, planificar, ejecutar, comparar y optimizar<sup>5</sup>.

Actualmente, en la realización de los ensayos clínicos existen tres elementos fundamentales a considerar: ciclos de tiempo (p. ej., tiempo requerido desde el inicio del ensayo hasta que se han reclutado todos los pacientes que requiere el protocolo), productividad (como número de enfermos incluidos por centro participante en el ensayo) y calidad (p. ej., evaluación de la exactitud y nivel de cumplimentación de la información registrada en los cuadernos de recogida de datos de los ensayos)<sup>6</sup>. Para conseguir el objetivo de completar los ensayos clínicos en el menor tiempo posible –manteniendo su calidad–, uno de los parámetros más importantes a tener en cuenta es la selección de investigadores y su eficiencia en la selección de pacientes.

El presente trabajo tiene como objetivos evaluar el grado de cumplimiento de los compromisos adquiridos por los investigadores y su eficiencia en la selección de enfermos para un ensayo clínico multinacional.

---

## Pacientes y métodos

---

El ensayo clínico escogido era un estudio de fase 2, aleatorizado, de rango de dosis (4 dosis del fármaco en investigación), controlado con placebo, de grupos paralelos, en pacientes con asma. El tamaño de los bloques de aleatorización era de cinco. El estudio comprendía un total de 7 visitas por paciente, distribuidas entre un período de selección simple ciego con placebo de 4 semanas (2 visitas) seguido, tras la aleatorización, de un período de tratamiento de 8 semanas. Los pacientes debían cumplir una serie de criterios de inclusión/exclusión tanto al comienzo de la fase de selección como en el momento de la aleatorización, y debían recoger diariamente una serie de parámetros en una tarjeta diseñada al efecto. El estudio pretendía aleatorizar a un mínimo de 550 pacientes, con un objetivo de 10 pacientes por centro. Para garantizar este mínimo, se estimó que el compromiso inicial necesario sería de 690 pacientes, asumiendo que no todos los centros cumplirían el compromiso mínimo de 10 pacientes por centro, hecho constatado en varios ensayos similares (en cuanto a diseño y población de pacientes) realizados con anterioridad por nuestra compañía en los países participantes. La selección de investigadores se realizó por el personal de los departamentos médicos de la compañía en cada uno de los países participantes, siguiendo un procedimiento normalizado de trabajo común a todos ellos. El compromiso del número de pacientes a aleatorizar por país fue responsabilidad de los departamentos médicos del promotor de cada país, una vez que el protocolo se había discutido con los investigadores potenciales. El período de reclutamiento estimado fue de 18 meses (de abril de 1994 a octubre de 1995). El estudio tenía previsto suspender la inclusión de enfermos cuando se alcanzase la cifra de 550 pacientes aleatorizados, y así se comunicó a los investigadores antes del

TABLA I  
NÚMERO DE ENFERMOS SELECCIONADOS Y ALEATORIZADOS POR PAÍS  
Y GRADO DE CUMPLIMIENTO DEL OBJETIVO PREVISTO

Pais (n.º de centros participantes)	Objetivo de pacientes a aleatorizar (n)	Pacientes seleccionados: n (mediana)	Pacientes aleatorizados: n (mediana)	Pacientes aleatorizados/ seleccionados (%)	Media (DE) de pacientes aleatorizados/ centro [mediana]	Pacientes aleatorizados frente a objetivo por país (%)
A (4)	25	63 (10,5)	13 (2,5)	20,6	3,2 (1,8) [2,5]	52
B (4)	60	21 (5)	13 (2,5)	61,9	3,2 (1,7) [2,5]	22
C (1)	10	7 (7)	7 (7)	100	7 (NA) [7]	70
D (10)	60	54 (3,5)	32 (1)	59,2	3,2 (2,4) [1]	53
E (6)	65	72 (13)	60 (11)	83,3	10 (1,8) [11]	92
F (11)	150	201 (14)	171 (10)	85	15,5 (3,2) [10]	114
G (11)	70	155 (13)	99 (10)	63,9	9 (2,6) [10]	141
H (5)	80	77 (16)	54 (10)	70,1	10,8 (1,1) [10]	68
I (9)	80	90 (8)	68 (7)	75,5	7,6 (1,9) [7]	85
J (3)	30	53 (12)	26 (8)	49	8,7 (0,3) [8]	87
K (5)	60	43 (7)	21 (5)	48,8	4,2 (6,5) [5]	35
Total (69)	690	836	564	67,5	8,2 (2,8) [7]	81,7

Pacientes seleccionados: pacientes incluidos en la fase de selección; pacientes aleatorizados: pacientes incluidos en la fase de tratamiento doble ciego. NA: no aplicable.

inicio del ensayo. El ensayo clínico se realizó siguiendo las normas de buena práctica clínica.

A través del programa informático de seguimiento utilizado en nuestra compañía, se recogieron los siguientes datos: fecha de comienzo (entrega de medicación) del primer y último centro por país, fecha del final del estudio (última visita del último paciente), número de pacientes seleccionados y número de pacientes aleatorizados previsto (objetivo) y real, por centro y por país. Se definieron los centros como “cumplidores” cuando cumplieron con el compromiso adquirido con el promotor del estudio, es decir, aleatorizaron un número de pacientes igual o superior a 10, y como “no cumplidores” cuando el número de pacientes aleatorizados fue inferior a 10. Se definieron como “mejores” los países con 2 o más centros cumplidores, y como “peores” aquellos en los que todos los centros fueron no cumplidores. Se definió como “eficiencia” en el reclutamiento de enfermos por los centros al porcentaje de pacientes aleatorizados respecto a seleccionados: un centro es más “eficiente” cuanto mayor es este porcentaje.

Se ha realizado un análisis descriptivo de los datos con cálculo de medias y medianas para las variables cuantitativas (pacientes aleatorizados por centro y por país) y porcentajes para la variable cualitativa (“eficiencia”). Se efectuaron las siguientes comparaciones de “eficiencia” mediante la prueba de la  $\chi^2$ : centros cumplidores frente a no cumplidores de todos los países, y la misma

limitada a los países “mejores”, así como de los centros no cumplidores de los países “mejores” frente a la “eficiencia” general de los países “peores”. Para dichas comparaciones se calculó asimismo, la *odds ratio* (OR) (con los intervalos de confianza [IC] del 95 % correspondientes) mediante el método de Mantel-Haenszel. Se compararon las siguientes variables cuantitativas (medias) mediante la prueba de ANOVA de una vía: pacientes aleatorizados por centro entre centros cumplidores y no cumplidores en los países, “mejores” y pacientes aleatorizados por centro entre países “mejores” y “peores”. El análisis estadístico se realizó mediante el programa EPIINFO 6.0 (CDC, Atlanta, GA, EE.UU.).

## Resultados

En el ensayo clínico participaron 69 hospitales en los siguientes países: Alemania, Bélgica, España, Francia, Holanda, Italia, Reino Unido, Suecia, Suiza, Australia y Sudáfrica. Diez de estos países iniciaron el ensayo entre abril y mayo de 1994; el país restante lo comenzó en septiembre de ese mismo año. En septiembre de 1995 (sólo unas semanas antes de la fecha inicialmente prevista) se suspendió la inclusión de enfermos en el estudio al haberse aleatorizado el paciente 550; en ese momento, se había seleccionado un 48 % más de enfermos (n = 836) de los que se aleatorizaron (n = 564). La tabla I recoge los resultados

TABLA II  
NÚMERO DE ENFERMOS SELECCIONADOS, ALEATORIZADOS Y EFICIENCIA  
DE LOS CENTROS CUMPLIDORES Y NO CUMPLIDORES POR PAÍS

País	N.º de centros		Pacientes seleccionados (n)		Pacientes aleatorizados (n)		Pacientes aleatorizados/seleccionados (%)	
	CC	CNC	CC	CNC	CC	CNC	CC	CNC
A	0	4	0	63	0	13	0	20,6
B	0	4	0	21	0	13	0	61,9
C	0	1	0	7	0	7	0	100
D	1	9	20	34	15	17	75	50
E	4	2	63	9	54	6	85,7	66,6
F	6	5	163	38	145	26	89,0	68,4
G	6	5	123	32	89	10	72,4	31,3
H	3	2	57	20	40	14	70,2	70,0
I	3	6	49	41	42	26	85,7	63,4
J	1	2	10	43	10	16	100	37,2
K	0	5	0	43	0	21	0	48,8
Total	24	45	485	351	395	169	81,4*	48,1*

\* $p < 0,001$ ; *odds ratio*, 4,7 (intervalo de confianza del 95 %, 3,4-6,5). Pacientes seleccionados: pacientes incluidos en fase de selección; pacientes aleatorizados: pacientes incluidos en fase de tratamiento doble ciego; CC: centros cumplidores ( $\geq 10$  pacientes aleatorizados); CNC: centros no cumplidores ( $< 10$  pacientes aleatorizados).

globales y por país sobre la selección de enfermos, cumplimiento del objetivo previsto y porcentaje de pacientes aleatorizados con respecto a seleccionados (“eficiencia”). Globalmente, el intervalo de pacientes seleccionados por investigador varió entre 1 y 44, y el de pacientes aleatorizados de 0 a 40. Sólo dos de los países cumplieron con el objetivo previsto. Diecisiete de los 69 centros (25 %) en 8 de los 11 países seleccionaron 58 pacientes, pero sólo aleatorizaron entre 0 y 1 por centro (7 en total); seis de estos investigadores pertenecían a 4 de los países “mejores” y 11 pertenecían a los 4 países “peores”.

En la tabla II se indica el rendimiento de los centros cumplidores y no cumplidores. Veinticuatro de los 69 investigadores (35 %) cumplieron el objetivo acordado. Estos centros se localizaron en 7 países, si bien sólo 5 de ellos tuvieron 2 o más centros de este tipo (países “mejores”). Los investigadores cumplidores ( $n = 24$ ) incluyeron en el estudio un 70 % (395 de 564) del total de pacientes aleatorizados, mientras los otros 45 aportaron el 30 % (169 de 564) restante. Los investigadores cumplidores fueron, además, significativamente más eficientes que los no cumplidores.

En la tabla III se exponen los resultados de los centros cumplidores y no cumplidores en los 5 países “mejores”. Los centros cumplidores en

dichos países aleatorizaron 370 pacientes de 455 seleccionados (82,1 %; intervalo: 10-40 por centro; mediana: 13,5), “eficiencia” que resultó significativamente superior ( $p < 0,001$ ) a la de los centros no cumplidores en esos mismos países (82 de 140; 58,6 %; intervalo: 0-9 por centro; mediana: 4). En estos 5 países se aleatorizaron un total de 452 pacientes (el 80,1 % del total) en 42 centros (media  $\pm$  DE de  $10,7 \pm 0,1$  pacientes por centro), mientras que en los 4 países “peores” se aleatorizaron 54 pacientes (9,6 %) en 14 centros ( $3,9 \pm 0,2$  enfermos por centro). El rendimiento medio fue 2,7 veces superior en los primeros ( $p < 0,001$ ). Los centros no cumplidores de los países “mejores” presentaron una “eficiencia” significativamente mejor ( $p < 0,01$ ) que la totalidad de centros de los 4 países “peores” (82 de 140; 58,6 % frente a 54 de 134; 40,3 %; *odds ratio* [OR], 2,0; IC del 95 %, 1,2-3,4).

Por último, hay que señalar que el 8 % (46 de 564) de los enfermos aleatorizados abandonaron el ensayo durante la fase de tratamiento: 23 de ellos (50 %) por acontecimientos adversos, y el resto por otras causas.

## Discusión

En general, la inclusión de enfermos en ensayos clínicos es un problema más importante de lo

TABLA III  
NÚMERO DE PACIENTES ALEATORIZADOS Y EFICIENCIA EN LOS CENTROS CUMPLIDORES Y NO CUMPLIDORES DE LOS PAÍSES "MEJORES" (AQUELLOS CON  $\geq 2$  CENTROS CUMPLIDORES)

Centros	n	Pacientes aleatorizados (n)	Media (DE) de pacientes aleatorizados/centro	Eficiencia: pacientes aleatorizados/seleccionados (%)
Cumplidores ( $\geq 10$ pacientes aleatorizados)	22	370	16,8 (2,8)*	370/455 (82,1)**
No cumplidores ( $< 10$ pacientes aleatorizados)	20	82	4,1 (3,1)*	82/140 (58,6)**

\* $p < 0,001$ ; \*\* $p < 0,001$ ; *odds ratio*, 3,0 (intervalo de confianza del 95 %, 2,7-4,7); pacientes seleccionados: pacientes incluidos en fase de selección; pacientes aleatorizados: pacientes incluidos en fase de tratamiento doble ciego.

que los propios investigadores piensan, y siempre resulta más difícil de lo que en un principio se planeó<sup>7,8</sup>. Los datos publicados indican que es infrecuente que un estudio reclute el número requerido de pacientes en el período de tiempo inicialmente previsto<sup>7,9-11</sup>. Una revisión de los estudios promovidos por los National Institutes of Health de EE.UU. demostró que en el 66 % de los ensayos multicéntricos nunca se llegó a seleccionar el número de enfermos previsto<sup>12</sup>. Así pues, en la mayor parte de los ensayos clínicos, muchos de los centros participantes son incapaces de seleccionar el número de enfermos según el calendario acordado; esto se debe a dos razones fundamentales: la selección de los investigadores participantes y las características del protocolo<sup>4</sup>. La selección correcta de los investigadores por parte de los promotores del estudio cobra, pues, una relevancia capital en su conducción. En este estudio se presentan algunos aspectos sobre el rendimiento y la selección de los centros en un estudio multicéntrico que no han sido descritos previamente, y que pueden ser de utilidad para quienes planifican la realización de este tipo de ensayos.

En nuestro estudio, sólo el 35 % de los investigadores cumplieron con el compromiso de aleatorizar, al menos, a 10 pacientes en el período de tiempo acordado; además, incluyeron el 70 % del total de enfermos requeridos por el ensayo. Las menores pérdidas prealeatorización de éstos (el 19 frente al 52 % en el resto de los centros) demuestran que, además, fueron significativamente más eficientes: los investigadores que más pacientes seleccionaron, lo hicieron con una mejor calidad. Hay que señalar, sin embargo, que el ensayo concluyó en el tiempo previsto gracias a la participación de centros no cumplidores.

Aunque desde la planificación del ensayo se previó que no todos los centros seleccionarían el número mínimo de pacientes previsto, el hecho de que la mayoría de los investigadores (2 de cada 3) no cumplieren con el objetivo acordado, y sobre todo que el 25 % sólo aleatorizaran entre 0 y 1 pacientes, indica que la selección de estos centros no se realizó correctamente. Otro hecho a destacar es que hubo centros no cumplidores en todos los países, incluso en aquellos considerados como "mejores", por lo que se puede afirmar que es un problema generalizado. Este es un asunto de difícil solución. Aunque las normas de buena práctica clínica de la Unión Europea especifican que es responsabilidad de los investigadores no sólo asegurar que disponen del personal, instalaciones y tiempo requeridos para completar el ensayo, sino que deben aportar datos retrospectivos de los pacientes que hubieran cumplido los criterios de selección de enfermos señalados en el protocolo, de forma que se pueda asegurar una adecuada tasa de selección<sup>13</sup>, hay que reconocer que esto último todavía no es una práctica común.

Aunque en la bibliografía se mencionan diversos métodos para mejorar la selección de enfermos en ensayos clínicos<sup>8,9,11</sup>, lo cierto es que son pocos los trabajos en los que se evalúa la utilidad de varios métodos de selección de pacientes en un mismo ensayo<sup>14</sup>. A pesar de esto, los promotores tenderán a minimizar o eliminar las contribuciones ineficientes en un ensayo como, por ejemplo, cancelando los centros que no han aleatorizado un número determinado de pacientes una vez transcurrido cierto tiempo desde el inicio del estudio. Ello puede conducir a una cierta reducción en el número de países y/o centros potencialmente candidatos a participar en ensayos futuros y, por consiguiente, a una mayor competencia entre ellos.



## Agradecimiento

Agradecemos al Dr. Antonio González la realización del análisis estadístico.

## BIBLIOGRAFÍA

1. Dal-Ré R. Coste, eficiencia y próximo futuro del desarrollo de nuevos fármacos. *Med Clin (Barc)* 1997; 108: 621-628.
2. DiMasi JA, Seibring MA, Lasagna L. New drug development in the United States from 1963 to 1992. *Clin Pharmacol Ther* 1994; 55: 609-622.
3. Drasdo AL, Halliday RG, Lumley CE, Walker SR. The strategy and management of successful global R&D. CMR 94-6R. Carshalton: Centre for Medicines Research, 1994.
4. Cutler S, Redmond C. Reducing drug development time –focus on patient recruitment. *Drug Inform J* 1995; 29 (Supl): 1.709-1.718.
5. Tassignon JP. Improving the drug development process. *Applied Clin Trials* 1995; 4: 20-32.
6. Barnett ST, James JA. Measuring the drug development process. *Applied Clin Trials* 1995; 4: 44-57.
7. Hunninghake DB, Darby CH, Probstfield JL. Recruitment experience in clinical trials: literature summary and annotated bibliography. *Controlled Clin Trials* 1987; 8 (Supl): 6-30.
8. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials (2.ª ed.). Littleton: PSG Publishing Co. Inc., 1985.
9. Spilker B, Cramer JA. Patient recruitment in clinical trials. Nueva York: Raven Press Ltd., 1992.
10. Agras WS, Bradford RH. Recruitment: an introduction. *Circulation* 1982; 66 (Supl 4): 2-5.
11. Meinert CL, Tonascia S. Clinical trials: design, conduct, and analysis. Nueva York: Oxford University Press, 1986.
12. Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomization. *Br Med J* 1984; 289: 1.281-1.284.
13. Commission of the European Communities. CPMP Working Party on Efficacy of Drugs. Good Clinical Practice for Trials on Medicinal Products in the E.C. III/3976/88-EN. Final. Bruselas, julio de 1990.
14. García-Losa M, Dal-Ré R, García-Méndez E. Improving patient recruitment in clinical trials: lessons from one multicentre study in asthma. *Br J Clin Pharmacol* 1995; 39: 211-212.

---

# La evaluación de la calidad de vida en el contexto del ensayo clínico

---

Xavier Badia<sup>a</sup> y Xavier Carné<sup>b</sup>

<sup>a</sup>Institut Universitari de Salut Pública de Catalunya. Universitat de Barcelona. L'Hospitalet de Llobregat.

<sup>b</sup>Servei de Farmacologia Clínica. Hospital Clínic i Provincial. Barcelona.

El concepto de calidad de vida asociada a la salud (CVAS) aparece en las sociedades occidentales en un momento en el que la esperanza de vida ha aumentado sensiblemente y en el que predomina la convicción de que el papel de la medicina no debe ser únicamente el de proporcionar muchos años de vida a las personas, sino, sobre todo, el de aportar una mejora en la calidad de los años vividos. En los últimos años, las publicaciones aparecidas en la prensa médica que tratan aspectos relacionados con la calidad de vida se han multiplicado. Si consultamos la bibliografía reciente relativa a la CVAS podremos observar que 3 ideas destacan sobre las demás. En primer lugar, la importancia del estado funcional del paciente (tanto físico, como social o mental) en el contexto de un concepto como el de calidad de vida, que se considera multidimensional<sup>1-5</sup>, en segundo lugar, el hecho de que la medición del estado de salud se considere un fenómeno innegablemente subjetivo<sup>6</sup>, y finalmente, la necesidad de obtener un valor numérico que represente la preferencia del enfermo por un determinado estado de salud<sup>7</sup>.

Dentro de la medicina, la evaluación de la CVAS se ha realizado en una gran variedad de contextos: *a)* en la práctica clínica; *b)* en la investigación de los servicios sanitarios; *c)* en la evaluación de nuevos medicamentos en ensayos clínicos; *d)* en el seguimiento de la salud de poblaciones y subgrupos de población, y *e)* en la asignación de los recursos sanitarios<sup>8</sup>.

Es fundamental reconocer que en el tratamiento de los pacientes en situación de práctica clínica, la percepción de un cambio en salud del propio enfermo es un importante indicador del éxito del tratamiento<sup>9</sup>. En cambio muchos ensayos clínicos se centran en evaluar el éxito del tratamiento utilizando variables de resultado como

el VEMS u otros parámetros fisiológicos. Sin negar la importancia de evaluar dichos parámetros, la inclusión del punto de vista del paciente debe contribuir de manera decisiva a evaluar los méritos relativos de los distintos tratamientos en los ensayos clínicos<sup>10</sup>.

En la presente revisión nos limitaremos a comentar las aportaciones de la evaluación de la CVAS al ensayo clínico.

## ¿Cuándo es importante evaluar la calidad de vida en un ensayo clínico?

---

En primer lugar debemos interrogarnos ¿cuándo la CVAS debe considerarse como un aspecto importante a evaluar en un ensayo clínico? De manera general, la CVAS parece particularmente importante evaluarla en aquellas enfermedades en las que las medidas habituales de resultado en los pacientes son de valor limitado. Por ejemplo, en el caso de la artritis reumatoide, donde la naturaleza de la enfermedad hace difícil la utilización de medidas que reflejen el resultado del tratamiento<sup>11</sup>. Sin embargo, también es importante evaluar la CVAS en otras situaciones donde el objetivo primario sea otro, como la supervivencia en el infarto de miocardio o si existe la necesidad de hacer balance entre la eficacia y los efectos adversos de los medicamentos antihipertensivos. Aunque las opiniones sobre cuándo evaluar la CVAS pueden ser muchas, su justificación parece mayor en 3 circunstancias clínicas concretas. En primer lugar, cuando la mejora en la calidad de vida es el principal resultado del tratamiento aplicado (p. ej., en los cuidados paliativos del cáncer o, en general, en pacientes con una enfermedad actualmente incurable). En segundo lugar, cuando 2 tratamientos son de una eficacia y una seguridad equivalentes, y el tratamiento nuevo puede aportar ventajas adicionales en la calidad de vida y, en tercer lugar, cuando éste ofrece ventajas menores en cuanto a la su-

TABLA I  
DIMENSIONES DE LA CALIDAD DE VIDA

<i>Dimensión</i>	<i>Ejemplos</i>
Función física	Movilidad, capacidad para desarrollar las actividades básicas de la vida diaria, desplazamiento
Función psicológica	Depresión, ansiedad, preocupación por el futuro
Función social	Pasatiempos, ocio, relaciones sociales, relaciones familiares
Función cognitiva	Memoria, alerta
Bienestar general	Salud general, percepciones de salud, satisfacción con la vida
Nivel económico	Ingresos económicos, trabajo

pervivencia del enfermo, a cambio de un evidente deterioro en su calidad de vida. También existen situaciones en las que la evaluación de la calidad de vida parece poco relevante. Por ejemplo, cuando examinamos un tratamiento potencialmente curativo con un perfil de efectos adversos muy favorable. Sin embargo, aun en estos casos algunos investigadores insisten en que como mínimo se deben considerar las posibles implicaciones en la calidad de vida y que, en cualquier caso, se debe justificar su no evaluación.

Se debe recordar que el análisis de la calidad de vida que aporta un determinado procedimiento o tratamiento forma parte del análisis de su efectividad o, lo que es igual, del estudio de los beneficios y riesgos atribuibles al mismo en las condiciones reales de la práctica. Para ello es conveniente que, con antelación, hayamos estudiado la eficacia del procedimiento, o sea, que conozcamos los beneficios y riesgos que aporta en las condiciones “ideales” de un experimento. Es por ello que el lugar “natural” de la medición de la calidad de vida a lo largo del desarrollo de un medicamento se encuentra en las fases avanzadas del mismo (fases III y IV), una vez tenemos una cierta idea del comportamiento “ideal” de la molécula<sup>12</sup>.

Existen dos posibles estrategias respecto al lugar que deben ocupar los estudios de calidad de vida en el contexto del ensayo clínico. La prime-

ra implica el realizar un estudio cuyo objetivo fundamental sea la medida de la CVAS, utilizando un diseño propio. La segunda estrategia supone la incorporación de la medida de la CVAS a un ensayo convencional, cuyo objetivo principal sea la valoración del beneficio-riesgo de la intervención, en el que se mida una variable “dura” en el contexto de un diseño bien conformado, o sea, aleatorizado, enmascarado y en paralelo. En nuestra opinión, es preferible esta segunda aproximación.

### ¿Qué instrumento de calidad de vida incluir en un ensayo clínico?

Los instrumentos de medida de la calidad de vida son cuestionarios que contienen las instrucciones para su cumplimentación y cuyo objetivo es la medición de las dimensiones de la CVAS (dolor, movilidad, etc.) a través de un número determinado de ítems descriptivos que ordenan los diferentes niveles de cada dimensión (“no tengo dolor”, “tengo dolor moderado”, “tengo mucho dolor”). Algunos ejemplos de dimensiones de la CVAS se exponen en la tabla I. La mayoría de medidas cumplen con las propiedades de validez (el instrumento mide aquello que pretende medir), fiabilidad en el sentido de repetibilidad (el instrumento produce los mismos resultados en mediciones repetidas bajo las mismas condiciones) y sensibilidad a los cambios (el instrumento produce diferentes resultados en mediciones repetidas, si las condiciones cambian)<sup>13</sup>. Esta última propiedad es esencial para incorporar la medida de la CVAS a los ensayos clínicos con medicamentos.

No existe un único instrumento de medida de la CVAS que pueda adaptarse a todas las áreas de estudio. Dada la variedad de ámbitos potenciales de aplicación parece poco probable que un único instrumento de medición sea capaz de satisfacer los requerimientos de todos los usuarios potenciales, en todas las circunstancias. En la práctica nos encontramos con una multitud de métodos y propuestas diferentes cuyo objetivo es el de medir la CVAS. Para un conocimiento en profundidad sobre el contenido de los instrumentos y de sus propiedades psicométricas los lectores pueden acudir a textos como los de McDowell y Newell<sup>14</sup> y el de Wilkin et al<sup>15</sup> y revisiones críticas como la de Anderson et al<sup>16</sup>.

Existen diferentes criterios a la hora de clasificar los instrumentos de medida de la CVAS, aunque la más aceptada sea la propuesta por Guyatt et al<sup>17</sup>, que distingue entre instrumentos genéricos e instrumentos específicos.

### *Instrumentos genéricos*

Los instrumentos genéricos no están relacionados con ningún tipo de enfermedad en particular y son aplicables tanto a la población general como a grupos específicos de pacientes, dado que incluyen un amplio espectro de dimensiones de la CVAS. Se pueden dividir en tres grandes grupos: las medidas de ítem único, los perfiles de salud y las medidas de utilidad o preferencia.

Las medidas de ítem único consisten normalmente en preguntar al paciente acerca de su salud o CVAS en un período temporal. Se trata de obtener la valoración de la salud general del enfermo. La pregunta más utilizada es del tipo: ¿cómo diría que se encuentra hoy de salud?, y el paciente debe responder en una escala ordinal que va de muy bien a muy mal<sup>18</sup>.

Un perfil de salud es un cuestionario que mide diferentes dimensiones de la CVAS. Éste proporciona una puntuación específica para cada uno de los aspectos de la CVAS que mide, y que, como mínimo, incluyen las dimensiones física, mental y social. Además, algunos cuestionarios producen una puntuación agregada de todos los aspectos, lo que se conoce como un índice. La mayoría de los perfiles de salud contienen expresiones que utiliza la gente cuando está, o cree estar, enferma, y son el fruto de muchos años de observación e investigación clínicas. Por su condición de genéricos, los perfiles de salud pueden utilizarse en cualquier ensayo clínico en que se considere importante evaluar la CVAS. Los perfiles de salud más utilizados son el *Sickness Impact Profile* (Perfil de las Consecuencias de la Enfermedad)<sup>19-22</sup>, el *Nottingham Health Profile*<sup>23,24</sup> (Perfil de Salud de Nottingham) y, recientemente, el *SF-36*<sup>25,26</sup>. Los tres han sido adaptados para su uso en España.

Los perfiles de salud ofrecen algunas ventajas en la evaluación de medicamentos. Por ejemplo, permiten determinar los efectos del tratamiento (incluso los no sospechados) en diferentes aspectos de la CVAS, sin necesidad de utilizar múltiples instrumentos. También pueden utilizarse para comparar diferentes tratamientos en enfermedades diferentes. Además, su validez y fiabilidad están generalmente muy bien estudiadas. La principal limitación de los mismos es que su contenido puede no adecuarse a aquellos aspectos de la CVAS más relevantes para la enfermedad en estudio y, en consecuencia, puede ser un instrumento poco sensible a los cambios antes y después del tratamiento.

Las medidas de utilidad se basan en las preferencias o utilidades que los individuos asignan a

los diferentes estados de salud del instrumento<sup>27</sup>. Las utilidades se obtienen por medio de diferentes técnicas de medida: la escala por categorías, el intercambio temporal y el juego estándar. Normalmente, la escala de medida tiene un recorrido que va de 0 (peor estado de salud imaginable, en ocasiones muerte) a 1 (mejor estado de salud imaginable). Los instrumentos estandarizados más utilizados son la matriz de Rosser y Kind<sup>28</sup>, la *Quality of Well Being Scale*<sup>29</sup> y el *EuroQoI*<sup>30</sup>. Únicamente este último está siendo adaptado para su uso en la población española<sup>31</sup>.

Las ventajas de la incorporación de medidas basadas en preferencias en un ensayo clínico incluyen la generabilidad, la capacidad de integrar mortalidad y morbilidad (especialmente útil si se realiza una evaluación económica paralela al ensayo clínico) y la incorporación de la preferencia temporal y el riesgo en los valores de utilidad. La mayor limitación de las medidas de utilidad en los ensayos clínicos consiste en la dificultad de administrar correctamente el instrumento y obtener preferencias de todos los pacientes que se están estudiando. Además, el hecho de que produzcan una puntuación agregada no permite estudiar cuál de los aspectos de la calidad de vida es el responsable del aumento o la disminución de la utilidad. Igual que los perfiles de salud, las medidas de utilidad pueden ser poco sensibles a los cambios pequeños pero clínicamente relevantes de los tratamientos.

### *Instrumentos específicos*

Los instrumentos específicos incluyen dimensiones de la CVAS de una determinada afección que se quiere estudiar. Las medidas concretas se centran en áreas en las que los clínicos, a través de la entrevista con el paciente y la exploración física, se encuentran más familiarizados. Normalmente incluyen preguntas sobre el impacto de los síntomas en dimensiones de la calidad de vida. Respecto a los instrumentos genéricos, tienen la ventaja de presentar una alta sensibilidad a los cambios ante el problema específico de salud que se está evaluando. Su mayor desventaja radica en que no permiten comparaciones entre diferentes enfermedades y, por lo tanto, no son útiles para conocer la eficacia relativa de diferentes tratamientos. En los últimos años se han desarrollado numerosos instrumentos específicos destinados a evaluar la calidad de vida de pacientes con artritis reumatoide, con síntomas gastrointestinales, oncológicos, sometidos a radioterapia de cuello y cabeza, en diálisis, epilépticos, con insuficiencia respiratoria crónica, en los ancianos y otros muchos más<sup>32</sup>.

## Consideraciones prácticas de la incorporación de medidas de la CVAS en los ensayos clínicos

La selección de instrumentos en el ensayo clínico requiere una cuidadosa estrategia, que es diferente en cada ensayo clínico, y que ha de tener en cuenta los propósitos potenciales del instrumento, es decir, su capacidad discriminante, evaluativa o predictiva. Normalmente, en el contexto del ensayo clínico, la capacidad evaluativa de una medida de CVAS será la más apreciada. La inclusión de instrumentos genéricos (perfil de salud y medida de utilidad) y de un instrumento específico en un ensayo clínico proporciona información de distintos aspectos de la CVAS, y es siempre aconsejable si los recursos lo permiten.

A continuación se comentan brevemente algunos de estos aspectos derivados de la incorporación de medidas de la CVAS en los ensayos clínicos: el número de pacientes necesario para el ensayo, el entrenamiento y monitorización de los entrevistadores, la duración y el seguimiento del ensayo, los efectos adversos, las variables de confusión y el efecto placebo<sup>34-36</sup>.

*Número de pacientes.* El cálculo del tamaño de la muestra de un estudio debe basarse en las diferencias que se consideren clínicamente significativas que debe aportar la intervención para que ésta pueda considerarse relevante. La magnitud de esta presunta diferencia (véase más adelante) es un valor "a priorístico" de carácter subjetivo y, por tanto, plenamente discutible, pero que debe basarse o bien en el consenso de expertos o, mejor aún, en estudios previos realizados *ad hoc*. En el caso de los instrumentos de CVAS estas condiciones no se cumplen en todos los casos. Además, hay que tener en cuenta que la experiencia en la utilización de medidas de la CVAS es insuficiente para determinar en muchos de los instrumentos y situaciones de aplicación la magnitud del cambio o diferencia importante de tratamiento.

*Entrenamiento y monitorización de los entrevistadores.* La inclusión de medidas de la CVAS en un ensayo clínico requiere, como es lógico, un esfuerzo adicional en cuanto al entrenamiento específico de los entrevistadores. Esta dificultad es máxima si el personal que debe realizar el trabajo es el propio personal sanitario, a menudo poco habituado a este tipo de cuestionarios. Por ello se precisa una metodología clara y común a todos los entrevistadores y estandarizada al inicio del proyecto. Se hace imprescindible la elaboración de un manual del entrevistador con pautas a se-

guir para solucionar los problemas y preguntas que vayan apareciendo a lo largo del ensayo. Además, en los ensayos clínicos de larga duración se deben programar sesiones de reentrenamiento y estandarización periódicas.

*Duración y seguimiento.* La mayoría de los ensayos clínicos en fase III son, relativamente, de corta duración, en comparación con los posibles cambios en la calidad de vida esperables. Algunas dimensiones de la CVAS experimentan una mejoría a largo plazo. Dichos cambios no se podrán detectar si no se sigue a los enfermos el tiempo suficiente. Valga como ejemplo el caso de un medicamento con capacidad para aumentar el hematócrito de los pacientes sometidos a diálisis. Es posible que dicho fármaco se asocie a un aumento de la función física de los sujetos a los 6 meses del tratamiento. En este contexto es esperable que, a medida que aumenta la función física del paciente, éste se encuentre mejor y, a la larga (quizá más de un año) es posible que mejore su función psicosocial. Es decir, dependiendo de los objetivos, se debe tener en cuenta el horizonte temporal adecuado que permita registrar todos los potenciales beneficios del medicamento.

*Acontecimientos adversos.* La falta de acontecimientos adversos de un determinado tratamiento no debe interpretarse como equivalente a calidad de vida. Existen 4 métodos de recogida de información clínica relativa a efectos adversos de los medicamentos: a) la recogida de información acerca de los fenómenos referidos espontáneamente por el paciente; b) el interrogatorio del participante mediante una pregunta estándar de tipo general; c) el interrogatorio utilizando un cuestionario de síntomas o mediante el uso de escalas visuales analógicas, y d) el interrogatorio utilizando una lista-guía cerrada<sup>12</sup>. Los sistemas abiertos (los dos primeros) se acostumbra a correlacionar con diferentes aspectos de la calidad de vida mejor que los dos últimos. Sin embargo, actualmente se considera que lo más recomendable es utilizar una combinación de diferentes métodos.

La inclusión en un cuestionario de ítems específicos asociados a un determinado efecto adverso de un medicamento puede sesgar la medición de la CVAS. Un ejemplo de ello podría ser el caso de la valoración de 3 hipotensores, un IECA, un bloqueador beta y un diurético, y en el que se plantea un cuestionario que incluye únicamente una medida que contiene aspectos relacionados con la disfunción que origina al sujeto el hecho de tener que orinar a menudo. Así

pues, las medidas de la CVAS tienen que estar orientadas a medir el impacto sobre la afección que se estudia y no sobre el medicamento.

*VARIABLES DE CONFUSIÓN.* La influencia de las variables externas que pueden actuar como potenciales variables de confusión debe analizarse siempre, incluso en el caso del estudio en el que dichas variables se distribuyen de forma homogénea entre el grupo de intervención y el de referencia. En concreto, los acontecimientos vitales (p. ej., muerte de un familiar, ganancias en el juego, etc.), pueden tener una influencia notable en la mejoría o el empeoramiento de un caso tratado, que, si no se tienen en cuenta, pueden atribuirse erróneamente al medicamento que se está ensayando.

*Efecto placebo.* Una mejoría en la CVAS se observa casi siempre en los participantes de un ensayo clínico. Generalmente el efecto placebo desaparece si el período de tratamiento es lo suficientemente largo (superior a 6 meses).

## **Metodología de la medida de la calidad de vida asociada a la salud**

Las propiedades psicométricas de un instrumento de medición, validez, fiabilidad y sensibilidad al cambio, deben ser siempre evaluadas y consideradas antes de utilizar un instrumento en un ensayo clínico<sup>37</sup>.

### *Validez*

El instrumento debe medir aquello que éste pretenda. Se han descrito diferentes tipos de validez<sup>38,39</sup> que se pueden resumir en tres: de criterio, de contenido y de concepto. La *validez de criterio* se refiere a la capacidad de un instrumento de medición para producir los mismos resultados que el “patrón de oro” o la medida de criterio. Debido a que en este caso no disponemos de un acuerdo en el “patrón de oro” para medir la CVAS, la validez se establece especificando las áreas o las dimensiones que deben medirse y las relaciones esperadas entre ellas y otras variables. La *validez de contenido* se refiere a la capacidad de un instrumento de medición de cubrir todas las dimensiones o áreas importantes de la CVAS (p. ej., funciones física, social, mental, percepciones, síntomas, etc.). La *validez de concepto* se refiere a la capacidad con que un instrumento de medida se relaciona con otras de manera que sea consistente con las hipótesis derivadas de los conceptos de la CVAS que se están

mediendo. Se analiza por medio de correlaciones entre el instrumento que se valida y otros instrumentos o variables que pretenden medir lo mismo.

El conocimiento de la situación para la que fue desarrollado un instrumento de medida es esencial para la evaluación de su validez. Un instrumento desarrollado para un propósito puede no ser apropiado para otros. Incluso el contenido concreto del instrumento dependerá de la pregunta específica planteada en la investigación. Si el objetivo es evaluar la necesidad de que se practique una determinada intervención a un paciente, éste debe ser amplio en contenido de dimensiones de la CVAS. En cambio, si se quiere evaluar un nuevo medicamento antiinflamatorio para la artrosis de rodilla, éste ha de ser específico, e incluir únicamente aquellas dimensiones de la CVAS afectadas por la artrosis de rodilla. En este caso, se supone que la mejoría de la CVAS en estas dimensiones dependerá de la eficacia del antiinflamatorio.

### *Fiabilidad*

La fiabilidad es el grado con que un instrumento de medición está libre de error aleatorio. Esta definición implica homogeneidad del contenido, es decir, los ítems del instrumento deben medir aspectos distintos del mismo concepto<sup>40</sup>.

Una segunda definición refiere que una medida de la CVAS es fiable (en el sentido de reproducible) cuando se obtiene el mismo resultado al aplicarlo (en dos ocasiones separadas por un intervalo de tiempo) a pacientes estables, es decir, a individuos que no han sufrido cambios en su estado de salud. A este tipo de fiabilidad se la denomina *test-retest* y es la que habitualmente se emplea cuando es el propio paciente el que rellena el cuestionario y no existen observadores de la medida.

Es especialmente importante evaluar la fiabilidad cuando se utilizan cuestionarios o escalas administrados y cumplimentados por un observador (normalmente un profesional sanitario). En este caso, la fiabilidad puede interpretarse como una medida de la asociación que examina el efecto de diferentes observadores en las puntuaciones de la prueba. Es la que se conoce más propiamente como *fiabilidad interobservador*. La evaluación de la misma se realiza mediante una prueba en la que diferentes observadores evalúan un estado de salud de un mismo sujeto. La *fiabilidad intraobservador* es la que mide la variación que ocurre en un mismo observador como resultado de la evaluación repetida del mismo estímulo.

En la evaluación de la fiabilidad de las medidas de la CVAS hay que demostrar tanto la homogeneidad de los cuestionarios (normalmente utilizando la prueba de K-D Richardson 20 o la prueba de alfa de Cronbach)<sup>41</sup>, como la fiabilidad test-retest (normalmente utilizando la correlación producto-momento de Pearson o el coeficiente de correlación intraclase)<sup>42</sup>, y como, en su caso, la fiabilidad intraobservador (mediante el coeficiente Kappa solo o ponderado)<sup>43</sup>.

Bullinger et al<sup>44</sup> proponen algunos criterios estadísticos para inferir la validez y la fiabilidad de una medida de la CVAS. En primer lugar, y según estos autores, la correlación de la puntuación de un ítem particular en relación con la puntuación total de la escala no ha de ser inferior a 0,40. Así mismo, la consistencia interna ha de ser superior a 0,70, y el coeficiente de correlación para demostrar la validez de la prueba ha de ser como mínimo de 0,50.

### *Sensibilidad a los cambios*

Como se comentó anteriormente, se trata de una propiedad fundamental en los ensayos clínicos. La sensibilidad a los cambios se define como la capacidad del instrumento para detectar cambios en la salud de los pacientes cuando éstos se producen. La sensibilidad al cambio es directamente proporcional al cambio en la puntuación cuando se produce una diferencia clínicamente relevante (véase el siguiente apartado), e inversamente proporcional a la variabilidad en la puntuación presente en los enfermos estables.

Por tanto, se necesitan dos tipos de estudios para evaluar la sensibilidad al cambio, uno en el que se examine la variabilidad en los pacientes estables y otro que demuestre que los cambios que se obtienen en la puntuación se corresponden a cambios reales en la salud de los sujetos<sup>45</sup>. En el primer caso, el cuestionario se administra a los mismos pacientes elegibles para el ensayo clínico durante un período de estabilidad de la enfermedad, en tantas ocasiones y con los mismos intervalos de tiempo que se han planeado en el ensayo. En el segundo estudio, se administra el cuestionario a los individuos antes y después de la aplicación de una intervención de conocida eficacia.

Como se verá más adelante, la sensibilidad al cambio se puede analizar mediante la prueba del tamaño del efecto<sup>46</sup>; o con la prueba propuesta por Guyatt et al<sup>47</sup>. Cuanto mayor sea la diferencia en la puntuación del cuestionario en los pacientes en los que existe un cambio real de salud, mayor será la sensibilidad del mismo. Por el con-

trario, cuanto mayor sea la diferencia en la puntuación obtenida en los sujetos estables, menor será la sensibilidad de la prueba.

### **El significado del efecto del tratamiento**

Uno de los principios fundamentales de la investigación clínica es aquel que establece que, tras el riguroso análisis estadístico de los resultados obtenidos en un ensayo, es imprescindible realizar una valoración de la relevancia o significación clínica de las diferencias observadas. Es importante, por tanto, distinguir entre significación estadística y clínica. El primero es un concepto ampliamente utilizado en medicina, en diseños basados en pruebas de hipótesis, y que nos proporciona una idea de la probabilidad de cometer un error de primera especie en el estudio que nos ocupa. Por otra parte no es un tema especialmente problemático y no será tratado en esta revisión. Sin embargo, en el contexto resbaladizo de los estudios de CVAS, sí que creemos que merece algún comentario el concepto de la significación clínica.

El problema podría plantearse en los siguientes términos: ¿cómo traducimos los cambios que se han observado en las puntuaciones de los instrumentos de calidad de vida en un ensayo clínico en términos clínicamente significativos? Como es bien sabido, la significación clínica se refiere a la importancia de la diferencia en los resultados clínicos observados entre un grupo de pacientes tratados y el grupo control, y se describe normalmente en términos de la magnitud de un resultado<sup>48</sup>. La significación clínica debe definirse en función de estándares externos obtenidos a partir del consenso de los expertos o, mejor, de los diferentes grupos de una comunidad interesados en el tema. Dichos estándares no son inmunes a las modas o al marketing de los laboratorios farmacéuticos. Así, por ejemplo, en el campo de la hipertensión arterial, en los últimos años los objetivos terapéuticos se han desplazado desde el de la protección cardiovascular (una variable relativamente "dura") a conceptos relacionados con la mejora de la calidad de vida. Sin embargo, hay que reconocer que el impacto de los diferentes tratamientos antihipertensivos sobre el bienestar del paciente es, generalmente, sutil. En este terreno, un instrumento de calidad de vida estandarizado y validado que cada vez se utiliza más es el Cuestionario de Bienestar Psicológico (CBP [*Psychological General Well-Being Index*])<sup>49,50</sup>. Dicho cuestionario tiene 6 dimensiones (bienestar positivo, autocontrol, vitalidad, salud general, depresión y ansiedad) y un total de 22 ítems. La

TABLA II  
DIFERENCIAS ENTRE DISTINTOS MEDICAMENTOS ANTIHIPERTENSIVOS EN EL  
CUESTIONARIO DE BIENESTAR PSICOLÓGICO SEGÚN DATOS DE 5 ENSAYOS CLÍNICOS

<i>Estudio</i>	<i>Medicamentos comparadores</i>	<i>Cambio medio en la puntuación</i>	<i>Porcentaje del cambio</i>
Croog et al (1986)	Captopril-metildopa	3,8**	3,5
	Captopril-propranolol	2,5*	2,3
	Propranolol-metildopa	1,0	1,0
Croog et al (1990)	Captopril-atenolol		
	Varones	1,0	0,9
	Mujeres	-0,8	-0,7
	Captopril-verapamilo		
	Varones	3,4	3,1
	Mujeres	1,0	0,9
Steiner et al (1990)	Atenolol-verapamilo		
	Varones	2,4	2,2
	Mujeres	1,8	1,6
Applegate et al (1991)	Captopril-propranolol	1,7	1,5
	Captopril-atenolol	0,3	0,3
	Enalapril-atenolol	0,7	0,5
	Enalapril-propranolol	2,1	1,9
Fletcher et al (1992)	Diltiazem-atenolol	3,1	2,8
	Diltiazem-enalapril	0,9	0,8
	Enalapril-atenolol	4,0	3,6
Fletcher et al (1992)	Cilazapril-atenolol	0,7	0,6
	Cilazapril-nifedipino	0,5	0,5

Adaptada de Hjermthal y Wiklund<sup>51</sup>; \*p < 0,05. \*\*p < 0,01.

puntuación va de 22 a 132; una mayor puntuación indica una percepción peor del bienestar. En diferentes ensayos clínicos que comparan el efecto sobre la calidad de vida de otros tantos antihipertensivos se observa que los cambios medios pre y postratamiento en el CBP son muy pequeños. En la tabla II se exponen dichos cambios; se aprecia una diferencia estadísticamente significativa entre el efecto del captopril y la metildopa en la CVAS, debida, probablemente, a los mayores efectos indeseados de esta última<sup>51</sup>. Sin embargo, es difícil conocer la significación clínica de estos cambios en la puntuación del cuestionario.

Otro problema es cómo considerar el impacto de los efectos indeseados sobre los diferentes aspectos de la calidad de vida. ¿Es posible comparar un impacto negativo sobre la actividad sexual de un medicamento antihipertensivo con un efecto negativo sobre la calidad del sueño, o del descanso, o sobre la vitalidad de otro medicamento? Esta pregunta es, por ahora, de difícil solución. En este sentido hay que decir que, la utilización de instrumentos que proporcionan una única puntuación agregada de múltiples dimensiones puede esconder efectos favorables o desfavorables de los medicamentos sobre las di-

mensiones de la calidad de vida por un efecto de compensación. Por ello es aconsejable proporcionar los resultados de los ensayos de forma agregada y, a la vez, por dimensiones (¡incluso por ítems!). De otro modo, la interpretación de la información es muy difícil.

#### *Diferencia en la puntuación clínicamente significativa*

En la práctica clínica, el juicio clínico de lo que supone una diferencia clínicamente significativa se basa en la experiencia clínica. La observación de cambios en la función y en el curso clínico en muchos pacientes que se corresponden con las variaciones en el resultado de un determinado test, nos proporciona la evidencia empírica necesaria para definir el juicio clínico. Sin embargo, en el campo de la investigación, la definición de la significación clínica presenta algunos problemas adicionales<sup>52</sup>. En primer lugar, por la falta de familiaridad por parte de los profesionales sanitarios con la idea de preguntar por la percepción de la salud en lugar de utilizar las habituales medidas "objetivas". Y en segundo lugar, porque un investigador no distingue fácilmente



entre la significación en la población y en el individuo.

Para facilitar su manejo, las puntuaciones de los instrumentos de calidad de vida con significado clínico para la población se pueden convertir en diferentes indicadores, como son los índices pronósticos, los índices de utilización de recursos o el coste de movilizar una unidad de resultado. Su utilización como índice pronóstico se ilustra a continuación utilizando como ejemplo los resultados del ensayo clínico antes mencionado (tabla II) con 3 antihipertensivos: captopril, metildopa y propranolol<sup>53</sup>. La interpretación que se podría hacer de los resultados de este ensayo es la siguiente: el 20 % de los pacientes tratados con metildopa y el 15 % de los tratados con propranolol hubieran permanecido igual o hubieran mejorado si se les hubiera tratado con captopril. Esto significa que de cada 500 hipertensos, 100 enfermos tratados con metildopa podrían haber mantenido su calidad de vida con captopril en lugar de experimentar un empeoramiento. De estos 100, a 45 se les podría haber ahorrado un empeoramiento sustancial en su bienestar, en su depresión o en su ansiedad. Este modelo ampliado a una población de un millón de hipertensos se traduce en 90.000 pacientes a los que se les ahorra una disminución sustancial de su bienestar general.

Un ejemplo de un índice de utilización de recursos podría ser la medición del estado de salud mental de un grupo de enfermos a través del Mental Health Inventory (MHI), y su asociación con la frecuentación de visitas médicas ambulatorias. De esta manera es posible calcular que el paciente mental que puntúa en el percentil más bajo de la distribución de puntuaciones del MHI, gasta tres veces más al año en cuidados de salud mental que el individuo medio situado en el percentil más alto<sup>54</sup>. En el último indicador propuesto, el coste de movilizar una unidad de resultado, la significación de los cambios en la calidad de vida se representan a partir del coste de pasar desde una puntuación X (peor salud) de un instrumento de medición de la CVAS a una puntuación Y (mejor salud).

Los clínicos necesitan conocer los cambios en las puntuaciones de los instrumentos de calidad de vida, no sólo para interpretar los cambios clínicos de los pacientes, sino también para determinar el tamaño de la muestra en un ensayo clínico, o para interpretar los estudios publicados en los que se incluyen medidas de calidad de vida. La definición del "mínimo cambio clínicamente importante" (MCCI) es esencial para la investigación clínica, dado que cambios muy pe-

queños pueden ser estadísticamente significativos sin tener ninguna relevancia clínica (especialmente en tamaños muestrales grandes). El MCCI se puede definir como la diferencia en la puntuación más pequeña en el dominio de interés que los pacientes perciben como beneficiosa y que obligaría a un cambio en el tratamiento del sujeto en ausencia de efectos indeseados graves o en presencia de un coste excesivo<sup>55</sup>.

Se han sugerido dos formas de interpretar el significado de los cambios en la calidad de vida que ya se han apuntado en los párrafos anteriores. La primera es la interpretación basada en la distribución estadística de las puntuaciones de un instrumento de medición de la calidad de vida en un estudio determinado, mediante la media y la desviación estándar de los cambios o las diferencias entre el grupo tratado y el control. La segunda es la interpretación basada en cambios clínicos externos<sup>56</sup>, en la que se comparan los cambios en las puntuaciones de los instrumentos de calidad de vida con los cambios clínicos externos o resultados.

#### *Interpretaciones de significación clínica basadas en la distribución estadística de las puntuaciones*

La medida más conocida es el tamaño del efecto, que compara la diferencia de la puntuación media pre y postratamiento con la variabilidad de sus puntuaciones pretratamiento (desviación estándar)<sup>46</sup>. El tamaño del efecto transforma los cambios pre y postratamiento en unidades estándar de medida que permiten la comparación de diferentes intervenciones. Cohen<sup>57</sup> estableció, arbitrariamente, que por debajo de 0,20 el efecto del tratamiento debe considerarse menor, entre 0,20 y 0,50 moderado, y con una cifra mayor de 0,80 importante. Sin embargo, aun definiendo estas categorías, existe la necesidad de hacerlas más interpretables, comparándolas con otros estándares clínicos o medidas sociales. Recientemente, Guyatt et al<sup>47</sup> proponen que la diferencia de la puntuación pre y postratamiento se divida por la desviación estándar de los individuos que han permanecido aparentemente estables (los que no han apreciado cambios en su salud). Esta última aproximación requiere, previamente al ensayo clínico, realizar estudios en pacientes estables clínicamente, que tengan el mismo problema de salud que los enfermos incluidos en el ensayo clínico.

Cuanto mayor sea el tamaño del efecto en los sujetos en los que exista un cambio real de salud, mayor será la sensibilidad del instrumento. Al

TABLA III  
RELACIÓN ENTRE EL NIVEL DE DEPENDENCIA EN LAS ACTIVIDADES DE LA VIDA DIARIA (AVD) SEGÚN LA PUNTUACIÓN TOTAL Y DIMENSIONES DEL PERFIL DE LAS CONSECUENCIAS DE LA ENFERMEDAD (PCE)

Número de AVD	N.º	Física <sup>a</sup>	Psicosocial <sup>b</sup>	Total PCE <sup>c</sup>
Independiente	4	22,1 (13,7)	3,4 (2,7)	12,8 (8,5)
Dependiente en una actividad	22	22,9 (13,8)	13,3 (14,4)	19,3 (12,3)
Dependiente en 2-3 actividades	13	32,2 (12,7)	17,4 (14,1)	27,5 (11,3)
Dependiente en 4-5 actividades	9	36,4 (5,4)	13,8 (2,9)	27,7 (5,1)
Dependiente en todas	16	42,6 (13,2)	20,8 (11,1)	33,5 (10,7)

ANOVA: <sup>a</sup>p < 0,001; <sup>b</sup>p = 0,1; <sup>c</sup>p < 0,001.

contrario, cuanto mayor sea la diferencia en la puntuación en los pacientes estables menor será su sensibilidad. Si no se conoce la sensibilidad a los cambios de un instrumento de medida de la CVAS utilizado en un ensayo clínico con resultados negativos, o bien el tratamiento no es eficaz o bien el instrumento no es sensible a los cambios.

#### *Interpretaciones de significación clínica basadas en los cambios clínicos externos o resultados*

En este caso, las puntuaciones de un cuestionario de medida de la calidad de vida se comparan con otras medidas clínicas o externas para conseguir una valoración de la significación clínica. Un método habitualmente utilizado es comparar la puntuación del cuestionario con una pregunta de evaluación general de la salud que contesta el médico, el paciente o ambos. La magnitud del efecto se puede conocer preguntando al paciente directamente por el cambio en su salud (en positivo o en negativo) que percibe como importante. Por ejemplo: desde la última vez que se visitó, ¿cómo diría que se encuentra usted hoy de su...? La escala de respuesta puede ir de muchísimo mejor a muchísimo peor. Esta metodología ha sido utilizada con escalas de 7 respuestas tipo Likert, confirmándose en varios estudios que la diferencia clínicamente importante se correspondería a un cambio medio en la puntuación de aproximadamente 0,5 en la pregunta de evaluación del cambio en la salud, siempre en contextos clínicos muy bien definidos<sup>58,59</sup>. Sin embargo, parece probable que la importancia clínica de la diferencia de un punto en la escala varíe según la entidad clínica o la historia natural de la enfermedad. Como extensión a este razonamiento algunos autores han sugerido la utilización de las curvas ROC (*receiver-operating characteristic curve*)<sup>60</sup>. Las curvas ROC ponen en relación las diferentes puntuaciones del cuestionario con la

probabilidad de que éstas representen un cambio verdadero o bien un falso positivo, en comparación con la pregunta de evaluación general de la salud. Los diferentes puntos de la curva sugieren puntos de corte que maximizan la sensibilidad o la especificidad en la detección de cambios, o que más discriminan entre pacientes que han mejorado o que han empeorado.

Otras maneras de obtener un valor de la significación clínica es comparando las puntuaciones obtenidas en los instrumentos de CVAS con otras variables externas. Por ejemplo, comparando los cambios en la puntuación de una medida de CVAS con los eventos comunes de la vida como una enfermedad grave, la pérdida del trabajo o la muerte de un ser querido. En 2 ensayos clínicos con antihipertensivos, Testa et al<sup>61</sup> compararon las puntuaciones del Cuestionario de Percepción de Salud General (CPSG) y la Escala de Eventos Vitales de Holmes y Rahe, y encontraron que un cambio en una unidad en la escala de percepción de salud general correspondía a 32 unidades de cambio en la escala de los eventos vitales. Así, cambios entre 0,10 y 0,20 unidades estandarizadas en el CPSG reflejan la influencia en la calidad de vida de eventos vitales importantes como, por ejemplo, la pérdida de un trabajo. Otra aproximación comparativa es la utilizada por Brook et al<sup>62</sup>, en la que una diferencia de 5 puntos en la Escala de Percepción de Salud es equivalente al efecto que el diagnóstico de hipertensión arterial produce sobre el paciente.

Finalmente, otro método propuesto para obtener una significación clínica de unos resultados es la comparación de las puntuaciones de los instrumentos de calidad de vida con algunas clasificaciones de gravedad clínica destinadas a conocer los pacientes con enfermedad leve o grave. En la tabla III se muestra la puntuación media de Perfil de las Consecuencias de la Enfermedad (*Sickness Impact Profile* [SIP]) para cada nivel de gravedad en pacientes parapléjicos y tetrapléjicos

cos. En la tabla III puede observarse que, a mayor gravedad, mayor puntuación en el SIP (Badia, observaciones no publicadas). Así, puntuaciones de 19 en el SIP reflejan que el paciente es dependiente en al menos una actividad básica de la vida diaria.

No se puede finalizar esta revisión sin reconocer que los clínicos sólo se sentirán cómodos manejando las puntuaciones de las medidas de calidad de vida cuando éstas les resulten tan familiares como la medición de la tensión arterial o la glucemia. La demostración de que los cambios en la puntuación de las medidas de la CVAS detectan cambios clínicamente importantes contribuirá, en gran medida, a su inclusión rutinaria en los ensayos clínicos con medicamentos<sup>63</sup>.

## BIBLIOGRAFÍA

1. Ware JE. Conceptualizing disease impact and treatment outcomes. *Cancer* 1984; 53 (Supl): 2.316-2.323.
2. Spitzer WO. State of science 1986: quality of life and functional status as target variables for research. *J Chron Dis* 1987; 40: 465-471.
3. Hörnquist JO. Quality of life: concept and assessment. *Scand J Soc Med* 1989; 18: 68-79.
4. Schroeder E. Concepts of health and illness. En: Cuyler JA, editor. *Health Indicators*. Nueva York: St Martin's Press, 1983.
5. World Health Organization. Constitution of the World Health Organization, Annex I. En: *Ten years of the World Health Organization*. Ginebra: WHO, 1958.
6. Guyatt GH, Feeny D, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; 118: 622-629.
7. Badia X. La medida de la calidad de vida relacionada con la salud en la evaluación económica: técnicas de obtención de preferencias de los estados de salud. En: Sacristán JA, Badia X, Rovira J, editores. *Farmacoeconomía: evaluación económica de medicamentos*. Madrid: Editores Médicos S.A., 1995; 56-71.
8. Patrick DL, Eriksson P. Health status and health policy. *Allocating resources to health care*. Nueva York: Oxford University press, 1993.
9. Deyo RA. The quality of life, research and care. *Ann Intern Med* 1991; 114: 695-696.
10. Pocock SJ. A perspective on the role of quality of life assessment in clinical trials. *Control Clin Trials* 1991; 12 (Supl): 257-265.
11. Bombardier C, Ware J, Russell IJ, Larson M, Chalmers A, Leighton Read J. Auranofin therapy and quality of life in patients with rheumatoid arthritis. *Am J Med* 1986; 81: 565-578.
12. Bakke O, Carné X, García-Alonso F. *Ensayos clínicos con medicamentos*. Barcelona: Mosby-Doyma, 1994.
13. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? *J Clin Epidemiol* 1992; 45: 1.341-1.345.
14. McDowell I, Newell C. *Measuring health. A guide to rating scales and questionnaires*. Nueva York: Oxford University Press, 1987.
15. Wilkin D, Hallan L, Doggett M. *Measures of need and outcome for primary health care*. Nueva York: Oxford Medical Publications, 1992.
16. Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health related quality of life. *Quality Life Res* 1993; 2: 369-395.
17. Guyatt GH, Veldhuyzen Van Zanten SJO, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: a taxonomy and review. *Can Med Assoc J* 1989; 140: 1.441-1.448.
18. Krause NM, Jay GM. What do global self-rated health items measure? *Med Care* 1994; 32: 930-942.
19. Bergner M, Bobbitt RA, Pollard WE, Martin DP, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981; 19: 787-805.
20. Badia X, Alonso J. Adaptación de una medida de la disfunción relacionada con la enfermedad: la versión española del Sickness Impact Profile. *Med Clin (Barc)* 1994; 102: 90-95.
21. Badia X, Alonso J. Re-scaling the Spanish version of the Sickness Impact Profile: an opportunity for the assessment of cross-cultural equivalence. *J Clin Epidemiol* 1995; 48: 949-957.
22. Badia X, Alonso J. Validity and reproducibility of the Spanish version of the Sickness Impact Profile. *J Clin Epidemiol* 1996; 49: 359-365.
23. Hunt SM, McEwen J, McKenna SP. *Measuring health status*. Londres: Croom Helm, 1986.
24. Alonso J, Antó JM, Moreno C. Spanish version of the Nottingham Health Profile: translation and preliminary validity. *Am J Public Health* 1990; 80: 704-708.
25. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992; 30: 473-483.
26. Alonso J, Prieto L, Antó JM. La versión española del SF-36 Health Survey (cuestionario de salud SF-36): un instrumento para la medida de los resultados clínicos. *Med Clin (Barc)* 1995; 104: 771-776.
27. Feeny DH, Torrance GW. Incorporating utility-based quality-of-life assessment measures in clinical trials. *Med Care* 1989; 27 (Supl): 190-204.
28. Rosser R, Kind P. A scale of valuations of states on illness: Is there a social consensus? *Int J Epidemiol* 1978; 7: 347-358.
29. Anderson JP, Bush JW, Berry CC. Classifying function for health outcome and quality of life evaluation. *Med Care* 1986; 24: 454-469.
30. EuroQol® Group. EuroQol® – a new facility for the measurement of health related quality of life. *Health Policy* 1990; 16: 199-208.

31. Badia X, Fernandez E, Segura A. Influence of sociodemographic and health status variables on valuation of health status in a Spanish population. *Eur J Public Health* 1995; 5: 87-93.
32. Spilker B, Molinek FR, Johnston KA, Simpson RL, Tilson MH. Quality of life bibliography indexes. *Med Care* 1990; 28 (Supl): 53-77.
33. Guyatt G, Jaeschke R. Measurement in clinical trials: choosing the appropriate approach. En: Spilker B, editor. *Quality of life assessments in clinical trials*. Nueva York: Raven Press, 1990; 37-46.
34. Guyatt GH, Feeny D, Patrick D. Issues in quality-of-life measurement in clinical trials. *Control Clin Trials* 1991; 12 (Supl): 81-90.
35. Cella DF, Wiklund I, Shumaker SA, Aaronson NK. Integrating health-related quality of life in to cross-national clinical trials. *Quality Life Res* 1993; 1 (Supl): 433-440.
36. Wiklund I, Dimenas E, Martine W. Factors of importance when evaluating quality of life in clinical trials. *Control Clin Trials* 1990; 11: 169-179.
37. Nunnally JC. *Psychometric theory*. Nueva York: McGraw-Hill, 1967.
38. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of wellbeing. *Health Serv Res* 1976; 11: 478-507.
39. Read JL, Quinn RJ, Hofer MA. Measuring overall health: an evaluation of three important approaches. *J Chron Dis* 1987; 40 (Supl): 75-79.
40. Streiner DL, Norman GR. Reliability. En: Steiner DL, Norman GR, editores. *Health measurement scales. A practical guide to their development and use*. Nueva York: Oxford University Press, 1989; 79-96.
41. Cronbach LJ. Coefficient alpha and the internal structure of the test. *Psychometrika* 1951; 16: 297-334.
42. Deyo RA, Diehr P, Patrick D. Reproducibility and responsiveness of health status measures. *Control Clin Trials* 1991; 12 (Supl): 142-158.
43. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement on partial credit. *Psychol Bull* 1968; 70: 213-220.
44. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Quality Life Res* 1994; 2: 451-459.
45. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *Can Med Assoc J* 1986; 134: 889-895.
46. Kazis LE, Jennifer JA, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27 (Supl): 178-189.
47. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1989; 40: 171-178.
48. Sackett DY, Haynes RB, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little Brown, 1985.
49. Dupuy HJ. The psychological general well-being (PGWB) index. En: Wenger NK, Mattson ME, Fuberg CP, editores. *Assessment of quality of life in clinical trials of cardiovascular therapies*. Nueva York: Le Jacq, 1984; 170-183.
50. Badia X, Gutierrez F, Wicklund I, Alonso J. Validity and reliability of the Spanish version of the "General Well-Being Index". *Quality Life Res* 1996; 5: 101-108.
51. Hjerdhal P, Wiklund IK. Quality of life on antihypertensive drug therapy: scientific end-point or marketing exercise? *J Hypertension* 1992; 10: 1.437-1.446.
52. Deyo R, Patrick DL. The significance of treatments effects: the clinical perspective. *Med Care* 1995; 33 (Supl): 286-291.
53. Croog SH, Levine S, Testa MA, Brown B, Bulpitt CJ, Jenkins D. The effects of antihypertensive therapy on the quality of life. *N Engl J Med* 1986; 314: 1.657-1.664.
54. Ware JE, Manning WG, Duan N. Health status and the use of outpatient mental health services. *Am Psychol* 1984; 39: 1.426-1.432.
55. Jaeschke R, Singer J, Guyatt G. Measurement of health status: ascertaining minimal clinically important difference. *Control Clin Trials* 1989; 10: 407-415.
56. Lydick E, Epstein RS. Interpretation of quality of life changes. *Quality Life Res* 1993; 2: 221-224.
57. Cohen J. *Statistical power analysis for the behavioral sciences*. Nueva York: Academic Press, 1977.
58. Juniper EF, Guyatt G, Epstein R, Ferrie PJ, Jaeschke R, Hille TK. Evaluation of impairment of health-related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992; 47: 76-83.
59. Juniper EF, Guyatt G, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994; 47: 81-87.
60. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis* 1986; 39: 897-906.
61. Testa MA, Anderson RB, Nackley JF, Hollenberg NK. Quality of life-hypertension study group. Quality of life and antihypertensive in men. A comparison of captopril and enalapril. *N Engl J Med* 1993; 328: 907-913.
62. Brook RH, Ware JE, Rogers WH. Does free care improve adults health?: results from a randomized controlled trial. *N Engl J Med* 1983; 309: 1.426-1.430.
63. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care* 1989; 27: 254-268.

Parte 2

---

**BIOESTADÍSTICA**

---

---

# La estadística en la investigación clínica de medicamentos. Estudio de artículos originales procedentes de centros españoles

---

M.C. Carré Llopis<sup>a</sup>, J. Jiménez Villa<sup>b</sup>, M. Martín Mateo<sup>c</sup>  
y F. Jané Carrencà<sup>d</sup>

<sup>a</sup>GDIS. Barcelona. <sup>b</sup>Regió Sanitària Costa de Ponent. Servei Català de la Salut. L'Hospitalet de Llobregat (Barcelona).

<sup>c</sup>Departament de Bioestadística. Facultat de Medicina. Universitat Autònoma de Barcelona. Bellaterra (Barcelona).

<sup>d</sup>Departament de Farmacologia i Psiquiatria. Universitat Autònoma de Barcelona. Hospital de la Santa Creu i Sant Pau. Barcelona.

El uso de la estadística en la investigación clínica de medicamentos es un fenómeno que se ha ido implantando progresivamente desde que se introdujo el método estadístico en el estudio de la eficacia y seguridad de los fármacos.

En uno de los primeros estudios sobre el uso de la estadística en los artículos originales, Schor y Karten<sup>1</sup> observaron que una gran proporción no usaban las pruebas estadísticas en el análisis de los resultados, o bien, contenían errores importantes. Si bien la complejidad del análisis estadístico ha aumentado desde entonces<sup>2-5</sup>, todavía se aprecian errores en un porcentaje importante de artículos<sup>6-8</sup>. En las publicaciones españolas este fenómeno es más evidente que en las anglosajonas<sup>9-11</sup>.

La mayor disponibilidad de paquetes informáticos ha contribuido a la extensión del uso de la estadística y a la utilización, y en ocasiones abuso, de técnicas sofisticadas. Esta tendencia no ha ido pareja a un aumento de su calidad y ha conducido a una disminución de la capacidad de lectura crítica de los profesionales sanitarios<sup>12</sup>.

Este estudio pretende cuantificar la frecuencia de utilización de las pruebas estadísticas y evaluar otros aspectos del análisis de los datos en los artículos originales sobre investigación clínica de medicamentos realizados en centros españoles.

## Material y métodos

---

Las unidades de estudio han sido los artículos de investigación clínica de medicamentos realizados en centros españoles. La población de estudio se obtuvo de la base de datos Embase. La estrategia de búsqueda se restringió al apartado de fármacos y productos químicos administrados exógenamente, a España, a estudios realizados

en seres humanos y a artículos publicados en 1975, 1980, 1985 y 1990.

Se consideró como un artículo de investigación a todo estudio que presentara datos originales, publicado en cualquier sección de la revista, con exclusión de las de cartas al director, editoriales, artículos de opinión o de actualización y revisiones. También se excluyeron los artículos que presentaban uno o más casos aislados de un proceso o enfermedad, o reacciones adversas por drogadicción.

Se incluyeron todos los artículos obtenidos en la búsqueda que cumplían los criterios de selección. Para la recogida de datos se utilizó únicamente la información que aparecía en el texto del artículo. Se diseñó un formulario que recogía datos sobre el análisis estadístico.

Se consideró que en un artículo se había usado la estadística cuando se había realizado algún tipo de análisis inferencial. En caso afirmativo, se valoró el grado de su descripción como "adecuado" cuando existía una descripción lo suficientemente detallada para que un lector con conocimientos de estadística pudiera comprender y reproducir el análisis estadístico utilizado. Se valoró como "inadecuado" cuando el análisis no se describía o se hacía de forma insuficiente, hecho que no permitía su replicabilidad.

Los procedimientos estadísticos se clasificaron según las categorías de Emerson y Colditz<sup>12</sup>. La clasificación original que constaba de 21 categorías ya había sido modificada por otros autores<sup>2-4</sup>. Dado que en esta revisión algunas de las categorías tenían muy pocos representantes, se redujeron a 14 (tabla I). En la categoría de medidas epidemiológicas se incluyó la de estandarización, que se refería a las medidas de incidencia y prevalencia<sup>12</sup>. Las categorías de tablas de vida, regresión de supervivencia y otros análisis de supervivencia se integraron en una sola. La categoría de potencia, que incluye la utilización de una diferencia detectable o útil en la determinación

TABLA I  
CATEGORÍAS DE PROCEDIMIENTOS ESTADÍSTICOS UTILIZADAS PARA EVALUAR  
EL CONTENIDO ESTADÍSTICO DE LOS ARTÍCULOS

<i>Categoría</i>	<i>Descripción</i>
Sin métodos estadísticos o sólo descriptivos	Sin contenido estadístico, o solamente métodos descriptivos (porcentaje, media, desviación estándar, error estándar...)
t de Student	t de Student para datos independientes y para datos apareados, Z de comparación de medias
Tabla de contingencia	Ji al cuadrado, prueba exacta de Fisher, prueba de McNemar
Pruebas no paramétricas	Prueba de Mann-Whitney, prueba de Wilcoxon, prueba de Kruskal-Wallis...
Medidas epidemiológicas	Riesgo relativo, <i>odds ratio</i> , sensibilidad, especificidad
Correlación de Pearson	r de Pearson
Regresión lineal simple	Regresión lineal con una variable dependiente y una independiente
Análisis de la variancia	Análisis de la variancia de una o varias vías, F de comparación de variancias
Correlación no paramétrica	Correlación de Spearman, tau de Kendall
Regresión lineal múltiple	Regresión lineal con más de una variable independiente
Comparaciones múltiples	Métodos de Bonferroni, Scheffé, Tukey, Duncan, Neumann-Keuls...
Tabla de vida y análisis de supervivencia	Métodos actuarial, de Kaplan-Meier, prueba log-rank, regresión de Cox, regresión logística
Otros procedimientos	Procedimientos no incluidos en las categorías anteriores (pruebas de Kolmogorov-Smirnov, análisis discriminante, índice kappa...)
Procedimiento no especificado	Utilización de un procedimiento estadístico sin identificación explícita

del tamaño de la muestra se eliminó, ya que se analizó como un aspecto específico del diseño del estudio. Se excluyeron las categorías de transformación de variables y de análisis coste-beneficio. En la categoría otros procedimientos, se incluyó la de análisis de la sensibilidad (sensibilidad a pequeños cambios en los parámetros del modelo o en otras asunciones). El resto de categorías no se modificaron. Se añadió una última categoría correspondiente a prueba estadística no identificada.

Cada procedimiento estadístico se contabilizó una única vez por artículo, independientemente del número de ocasiones en que se mencionaba. Cada artículo se contabilizó una sola vez en cada categoría, aunque utilizara más de un procedi-

miento incluido en la misma. Dado que en un mismo trabajo podían coexistir procedimientos de diferentes categorías, un artículo podía estar contabilizado en más de una de ellas.

Emerson y Colditz denominaron accesibilidad estadística al porcentaje de artículos que una persona podría leer (y comprender) con determinados conocimientos<sup>12</sup>. Para poder construir este índice se ordenaron los procedimientos según su complejidad, tal como se presenta en la tabla I, de forma que cada artículo se codificó en la categoría más elevada. En su determinación se excluyeron los artículos que utilizaban pruebas no identificadas.

La presentación de los resultados en el artículo se consideró "adecuada" cuando era fácil-

TABLA II  
ARTÍCULOS INCLUIDOS EN EL ESTUDIO SEGÚN LA NACIONALIDAD  
Y EL AÑO DE PUBLICACIÓN DE LAS REVISTAS

Nacionalidad	Año			Total
	1980	1985	1990	
Española	63 (78,8)	41 (64,1)	73 (50,7)	177 (61,5)
Extranjera	17 (21,3)	23 (35,9)	71 (49,3)	111 (38,5)
Total	80 (27,8)	64 (22,2)	144 (50)	288 (100)

$\chi^2 = 0,001$ . Resultados expresados como n.º de artículos (tanto por ciento).

mente comprensible y reflejaba suficientemente la estrategia de análisis y “no adecuada” cuando existían errores o críticas tales como falta de claridad o confusión en los datos presentados, o si éstos aparecían en una secuencia que no permitía seguir el análisis. Se registró si se utilizaban intervalos de confianza (IC) para indicar la magnitud del resultado principal. En los estudios comparativos se recogió la existencia de significación estadística y, en caso afirmativo, al grupo al que favorecía. En los estudios con resultados estadísticamente no significativos (estudios negativos), se registró si se había calculado la potencia estadística o, como mínimo, si se discutía explícitamente la posibilidad de incurrir en un error beta o de que el tamaño de la muestra hubiera sido insuficiente para detectar el efecto de interés.

La recogida de datos se realizó por un solo observador, que procedió a la lectura del artículo identificando la existencia y el grado de cumplimiento de cada uno de los ítems. Previamente a la recogida de datos se realizó un análisis de concordancia entre dos observadores de forma enmascarada, con 30 artículos de investigación clínica de medicamentos indizados en Embase en años distintos a los revisados. Se obtuvo un índice kappa que variaba, según los ítems, entre 0,82 y 0,93.

Para el análisis, las revistas se consideraron españolas cuando eran editadas en España, aunque lo fueran en inglés, y extranjeras en caso contrario.

El análisis estadístico se efectuó con el programa SPSS/PC + 4.0. La comparación entre las revistas españolas y las extranjeras se realizó mediante la prueba de la ji al cuadrado o la prueba exacta de Fisher cuando se trataba de proporciones y la prueba de la t de Student-Fisher cuando eran medias. La comparación entre los años de publicación se realizó mediante la prueba de tendencia lineal de Mantel para las proporciones y el análisis de la variancia para las me-

dias. En todos los contrastes de hipótesis, el nivel de significación se fijó en el 5 %.

## Resultados

Se incluyeron 288 artículos de investigación clínica de medicamentos realizados en centros españoles. En 1975 no se halló ninguno, mientras que la mitad correspondían a 1990. Ciento setenta y siete artículos (61,5 %) se publicaron en revistas españolas y 111 (38,5 %) en revistas extranjeras. El porcentaje de artículos en revistas extranjeras aumentó significativamente en los años evaluados (tabla II).

En 211 artículos (73,3 %; IC 95 %: 68,2-78,4) se utilizaba algún procedimiento estadístico, con un aumento significativo ( $p < 0,001$ ) a lo largo de los años revisados (tabla III). Dicho porcentaje fue superior en las revistas extranjeras que en las españolas (84,7 frente al 66,1 %,  $p < 0,001$ ).

En 121 artículos (57,3 %; IC 95 %: 50,6-64,0) se describían de forma adecuada las pruebas estadísticas usadas en el análisis de resultados y no se observaron diferencias estadísticamente significativas entre revistas españolas y extranjeras ( $p = 0,25$ ), aunque se aprecia una mejora significativa entre los años estudiados ( $p < 0,001$ ), principalmente entre 1980 (27,3 %) y 1985 (58 %) (tabla III).

La categoría utilizada con mayor frecuencia fue la de la t de Student (33,7 %; IC 95 %: 28,2-39,2), seguida de la de tablas de contingencia (28,8 %; IC 95 %: 23,6-34,0) y de la falta de uso de la estadística o sólo de métodos descriptivos (27,8 %; IC 95 %: 22,6-33,0) (tabla IV). En relación con las revistas españolas, en las extranjeras utilizaron con más frecuencia las categorías de tablas de contingencia (43,2 %; IC 95 %: 37,5-48,9), análisis de supervivencia (18,9 %, IC 95 %: 14,4-23,4) y otros procedimientos (4,5 %; IC 95 %: 2,1-6,9). En cambio, en las revistas españolas fue más frecuente la categoría de falta de uso de estadística o sólo de



TABLA III  
USO DE PROCEDIMIENTOS ESTADÍSTICOS Y SU GRADO DE DESCRIPCIÓN  
EN EL ARTÍCULO EN EL ANÁLISIS DE RESULTADOS, SEGÚN LA NACIONALIDAD  
Y AÑO DE PUBLICACIÓN DE LA REVISTA

	Nacionalidad			Año				Total (n = 288)
	Española (n = 177)	Extranjera (n = 111)	p	1980 (n = 80)	1985 (n = 64)	1990 (n = 144)	p	
Usan estadística	117 (66,1)	94 (84,7)	< 0,001	44 (55,0)	50 (76,9)	117 (81,8)	< 0,001	211 (73,3)
Descripción adecuada de la estadística <sup>a</sup>	63 (53,4)	58 (61,7)	0,25	12 (27,3)	29 (58,0)	80 (68,4)	< 0,001	121 (57,3)

<sup>a</sup>En estudios que usan estadística.

Resultados expresados como n.º de artículos (tanto por ciento).

procedimientos descriptivos (35,6 %, IC 95 %: 30,1-41,1). El 8,7 % (IC 95 %: 5,4-12,0) de artículos no identificaban las pruebas estadísticas utilizadas, sin observarse diferencias entre revistas españolas y extranjeras en este aspecto.

En los años estudiados se observó un aumento estadísticamente significativo en la utilización de las categorías de t de Student, tablas de contingencia, pruebas no paramétricas, análisis de la variancia, comparaciones múltiples y análisis de supervivencia (tabla IV). Tan sólo se observó un descenso estadísticamente significativo en la categoría de falta de uso de estadística o sólo de técnicas descriptivas.

La complejidad del análisis estadístico fue superior en las revistas extranjeras. Mientras que en las revistas españolas con el conocimiento de las siete primeras categorías se podía comprender el 85 % de los artículos, en las extranjeras ello ocurría únicamente con el 60,6 % (tabla V). La complejidad también aumentó con los años valorados. El conocimiento de las pruebas incluidas en las siete primeras categorías permitía comprender el 95,7 % de los artículos publicados en 1980, el 80,4 % en 1985 y únicamente el 63,2 % en 1990. Cabe destacar que en 1980 el conocimiento de las tablas de contingencia y de la t de Student permitía comprender el 90 % de los ar-

TABLA IV  
PRUEBAS ESTADÍSTICAS UTILIZADAS SEGÚN LA NACIONALIDAD Y AÑO  
DE PUBLICACIÓN DE LA REVISTA

Categoría	Nacionalidad			Año				Total (n = 288)
	Española (n = 177)	Extranjera (n = 111)	p	1980 (n = 80)	1985 (n = 64)	1990 (n = 144)	p	
No estadística o sólo descriptiva	63 (35,6)	17 (15,3)	0,0002	37 (46,3)	15 (23,1)	28 (19,6)	< 0,0000	80 (27,8)
t de Student	61 (34,5)	36 (32,4)	0,72	18 (22,5)	23 (35,4)	56 (39,2)	0,04	97 (33,7)
Tablas de contingencia	35 (19,8)	48 (43,2)	< 0,0000	11 (13,8)	16 (24,6)	56 (39,6)	0,0002	83 (28,8)
Pruebas no paramétricas	15 (8,5)	11 (9,9)	0,68	1 (1,3)	8 (12,3)	17 (11,9)	0,017	26 (9,0)
Medidas epidemiológicas	-	2 (1,8)	0,08	-	-	2 (1,4)	0,36	2 (0,7)
Correlación de Pearson	6 (3,4)	3 (2,7)	0,74	3 (3,8)	1 (1,5)	5 (3,5)	0,70	9 (3,1)
Regresión lineal simple	8 (4,5)	2 (1,8)	0,22	2 (2,5)	4 (6,2)	4 (2,8)	0,40	10 (3,5)
Análisis de la variancia	13 (7,3)	14 (12,6)	0,14	-	5 (7,7)	22 (15,4)	0,0007	27 (9,4)
Correlación no paramétrica	-	1 (0,9)	0,21	-	1 (1,5)	-	0,18	1 (0,3)
Regresión lineal múltiple	2 (1,1)	2 (1,8)	0,64	-	1 (1,5)	3 (2,1)	0,44	4 (1,4)
Comparaciones múltiples	5 (2,8)	4 (3,6)	0,71-	-	-	9 (6,3)	0,009	9 (3,1)
Análisis de supervivencia	9 (5,1)	21 (18,9)	0,0002	2 (2,5)	3 (4,6)	25 (17,5)	0,0005	30 (10,4)
Otros procedimientos	1 (0,6)	5 (4,5)	0,02	-	2 (3,1)	4 (2,8)	0,3	6 (2,1)
Prueba no identificada	17 (7,2)	12 (6,7)	0,88	10 (11,9)	8 (9,0)	11 (4,5)	0,32	29 (7,1)
Total de pruebas utilizadas	235	178		84	89	242		413

Resultados expresados como n.º de artículos (tanto por ciento).

TABLA V  
ACCESIBILIDAD ESTADÍSTICA SEGÚN LA NACIONALIDAD Y AÑO DE PUBLICACIÓN  
DE LA REVISTA

Categoría	Nacionalidad		Año			Total (n = 259)
	Española (n = 160)	Extranjera (n = 99)	1980 (n = 70)	1985 (n = 56)	1990 (n = 133)	
No estadística o sólo descriptiva	61 (38,1)	16 (16,1)	37 (52,9)	13 (23,2)	27 (20,3)	77 (29,7)
t de Student	28 (55,6)	12 (28,3)	13 (71,4)	13 (46,4)	14 (30,8)	40 (45,2)
Tablas de contingencia	29 (73,8)	23 (51,5)	13 (90,0)	13 (69,6)	26 (50,4)	52 (65,3)
Pruebas no paramétricas	12 (81,3)	7 (58,6)	- (90,0)	5 (78,6)	14 (60,9)	19 (72,6)
Medidas epidemiológicas	- (81,3)	- (58,6)	- (90,0)	- (78,6)	- (60,9)	- (72,6)
Correlación de Pearson	2 (82,5)	2 (60,6)	2 (92,9)	- (78,6)	2 (62,4)	4 (74,1)
Regresión lineal simple	4 (85,0)	- (60,6)	2 (95,7)	1 (80,4)	1 (63,2)	4 (75,7)
Análisis de la variancia	9 (90,6)	10 (70,7)	- (95,7)	6 (91,1)	13 (72,9)	19 (83,0)
Correlación no paramétrica	- (90,6)	- (70,0)	- (95,7)	- (91,1)	- (72,9)	- (83,0)
Regresión lineal múltiple	1 (91,3)	- (70,0)	- (95,7)	- (91,1)	1 (73,7)	1 (83,4)
Comparaciones múltiples	4 (93,8)	3 (73,7)	- (95,7)	- (91,1)	7 (78,9)	7 (86,1)
Análisis de supervivencia	9 (99,4)	21 (94,9)	3 (100)	3 (96,4)	24 (97,0)	30 (97,7)
Otros procedimientos	1 (100)	5 (100)	- (100)	2 (100)	4 (100)	6 (100)

<sup>a</sup>Se excluyeron los artículos que utilizaban pruebas estadísticas no identificadas.

Resultados expresados como n° de artículos (tanto por ciento).

tículos publicados, mientras que sólo permitía el acceso al 50,4 % de los publicados en 1990.

El 49 % (IC 95 %: 43,2-54,8) de los artículos presentaban de forma adecuada los resultados. Se ha observado una mejoría estadísticamente significativa en los años estudiados (tabla VI). Sólo el 22,2 % (IC 95 %: 17,4-27,0) de los artículos revisados utilizaban IC, porcentaje que fue superior en las revistas extranjeras (32,4 frente al 15,8 %,  $p = 0,002$ ) y que presentó una mejoría significativa en los años estudiados, especialmente entre 1980 y 1985 (5 y 28,1 %, respectivamente). El 46,6 % (IC 95 %: 40,8-52,4) de los artículos publicados no encontraron diferencias estadísticamente significativas, mientras que en el 49,1 % de los casos las diferencias favorecían al grupo al que se administró la terapia en estudio. Estos porcentajes no mostraron diferencias significativas en función de la nacionalidad de las revistas ni de los años estudiados. De los 74 estudios negativos, únicamente 7 (9,5 %; IC 95 %: 6,1-12,9) habían calculado o discutido la potencia del estudio. Se ha observado una tendencia estadísticamente significativa a la disminución de dicho porcentaje en los años estudiados.

## Discusión

En este estudio se evaluó la descripción de los procedimientos estadísticos utilizados que constan en el artículo. No se pretendía evaluar la ade-

cuación de las pruebas, ni el cumplimiento de las condiciones para su aplicación.

Se utilizó Embase como base de datos, ya que desde sus inicios ha utilizado al país como descriptor, su cobertura está focalizada en las revistas de farmacología y europeas, de las que indiza todos los artículos sobre fármacos. Sin embargo, evidentemente, los resultados de este estudio no reflejan exhaustivamente la situación española en la investigación de medicamentos, ya que no todos los estudios son publicados ni todos los que lo son se incluyen en las bases de datos bibliográficas.

De forma semejante a los estudios de Feinstein<sup>6</sup> y Pocock et al<sup>13</sup>, los artículos han sido revisados por un único evaluador que, aunque está avezado en la lectura crítica de la bibliografía biomédica, podría haber introducido un sesgo, especialmente en los ítems que requerían una cierta subjetividad en su evaluación. Sin embargo, se observaron concordancias elevadas con un segundo observador "ciego" antes del inicio del estudio. Por otro lado, la evaluación por un único revisor podría disminuir la variabilidad de los resultados.

En un estudio realizado entre 1966 y 1969, Lionel y Herxheimer<sup>14</sup> observaron que sólo se utilizaba la estadística en el 30 % de los artículos. Este porcentaje ha aumentado y actualmente más del 75 % de las publicaciones usan algún tipo de estadística inferencial<sup>5,15-22</sup>, porcentaje que, en al-

TABLA VI  
USO DE TÉCNICAS ESTADÍSTICAS EN LA PRESENTACIÓN DE RESULTADOS  
SEGÚN LA NACIONALIDAD DE LA REVISTA

	Nacionalidad			Año			p	Total (n = 288)
	Española (n = 177)	Extranjera (n = 111)	p	1980 (n = 80)	1985 (n = 64)	1990 (n = 144)		
Presentación adecuada de resultados	82 (46,3)	59 (53,1)	< 0,001	32 (40,0)	33 (51,6)	76 (52,8)	< 0,000	141 (49,0)
Uso de intervalos de confianza	28 (15,8)	36 (32,4)	0,002	4 (5,0)	18 (28,1)	42 (29,2)	0,002	64 (22,2)
Grupo favorecido por las diferencias <sup>a</sup>			0,49				0,88	
Intervención	43 (53,0)	36 (45,0)		15 (51,7)	17 (47,2)	46 (47,9)		79 (49,1)
Control	4 (4,9)	3 (3,8)		2 (6,9)	2 (5,6)	3 (3,2)		7 (4,3)
Resultados sin diferencias significativas <sup>a</sup>	34 (42,0)	41 (51,3)		12 (41,3)	17 (47,2)	47 (50,0)		75 (46,6)
Total de estudios comparativos	81	80		29	36	96		161
Cálculo de la potencia <sup>b</sup>	5 (14,7)	2 (4,9)	0,15	4 (33,3)	2 (11,8)	1 (2,1)	0,009	7 (9,3)

<sup>a</sup>Sobre estudios comparativos. <sup>b</sup>Sobre estudios negativos.

Resultados expresados como n° de artículos (tanto por ciento).

gunos casos, llega al 90 %<sup>23</sup>. En este estudio, se ha observado que el 73,3 % de los artículos utilizaban la estadística, si bien, esta cifra era significativamente inferior en las revistas españolas (66,1 %) frente a las extranjeras (84,7 %).

El porcentaje de artículos en que la descripción del análisis estadístico puede considerarse adecuada varía ampliamente entre los distintos estudios y se ha observado en el 6 %<sup>24</sup>, el 26 %<sup>25</sup>, del 70 al 75 %<sup>13</sup>, el 85 %<sup>23</sup>, o hasta el 92,9 %<sup>26</sup>. En este estudio se ha considerado correcta en el 57,3 % de los casos, cifra muy similar al 60 % observado por Evans y Pollock<sup>27</sup>. Este porcentaje ha ido mejorando progresivamente en los años evaluados, aunque sin apreciarse diferencias entre revistas españolas y extranjeras.

Existe una gran diversidad de pruebas estadísticas que pueden utilizarse en investigación clínica, dependiendo del objetivo, del diseño del estudio y de las variables implicadas. Esta diversidad hace que la frecuencia de su uso varíe ampliamente en función de los temas de interés o de la especialidad médica<sup>2-4,16</sup>. En este estudio se aprecia que todavía el 8,7 % de los artículos no identifican las pruebas estadísticas que utilizan, semejante al 8,1 % observado por Jiménez Villa et al<sup>26</sup> y que el porcentaje de artículos que usan procedimientos estadísticos más allá de los puramente descriptivos es únicamente del 72,2 %. Sin embargo, se observa un incremento hasta el 80,4 % en 1990, porcentaje similar al 81,1 % observado en *Medicina Clínica* en 1991 y 1992<sup>28</sup> o

en *Atención Primaria* en 1990<sup>26</sup>. Las pruebas estadísticas más utilizadas en este estudio son las bivariantes, la t de Student (33,7 %) y las tablas de contingencia (28,8 %). Los porcentajes observados en 1990 son similares a los encontrados en *Medicina Clínica* en 1991 y 1992, con el 33,3 y el 49,2 %, respectivamente y al 29 % observado en la t de Student en *Atención Primaria* en 1990<sup>26</sup>, aunque en este mismo estudio se observa el 69,4 % de pruebas de contingencia. Probablemente las pequeñas diferencias observadas en relación con el estudio de Mora et al<sup>28</sup> y de Jiménez Villa et al<sup>26</sup> se deban a que únicamente se han evaluado artículos de investigación de medicamentos.

En este estudio, los procedimientos estadísticos se han agrupado en categorías y se han ordenado siguiendo las recomendaciones de Emerson y Colditz. Aunque puede ser discutible, ya que incluye cierto grado de subjetividad, estos autores justifican dicha ordenación por la cadencia y el nivel de adquisición de conocimientos estadísticos de un alumno que inicia su formación partiendo de cero, de forma que en las siete primeras categorías se incluyen los métodos bivariantes que corresponden a los contenidos tradicionales de un curso de estadística básica. Además, el índice de accesibilidad permite comparar el nivel de complejidad de las pruebas estadísticas utilizadas.

Aunque sólo se sobrepasa el nivel de la estadística bivariante en un escaso porcentaje de ar-

tículos, el grado de complejidad es superior al observado por Emerson y Colditz. Estos autores determinaron una accesibilidad del 73 % para las tres primeras categorías y del 83 % para las siete primeras<sup>12</sup>, mientras que en ese estudio se han observado del 65,3 y del 75,7 %, respectivamente. Sin embargo, la accesibilidad habría disminuido en 1990 hasta el 50,4 y el 63,3 % para las tres y las siete primeras categorías respectivamente, porcentajes que se asemejan al 40,2 y al 61,7 % observados por Mora et al<sup>28</sup> en *Medicina Clínica* entre 1991 y 1992. Estas diferencias podrían deberse al distinto período de estudio, ya que Emerson y Colditz evaluaron artículos publicados en 1978 y 1979 y Mora et al en 1991 y 1992.

En este estudio se ha observado que, globalmente, la accesibilidad estadística es menor en los artículos procedentes de revistas extranjeras, de acuerdo con los hallazgos de otros autores<sup>17,28</sup>. El uso de pruebas estadísticas más allá de las descriptivas es del 83,9 % en las revistas extranjeras, superior al 60 % observado en la revista *The New England Journal of Medicine* en 1878-1879<sup>12</sup> y al 70 % de *Atención Primaria* en 1989-1991<sup>10</sup>, y similar al 88-89 % observado en un estudio de diversas revistas médicas en la década de los ochenta<sup>16</sup>.

El índice de accesibilidad ha ido disminuyendo progresivamente en los años estudiados, tanto en las revistas extranjeras como en las españolas, aunque puede apreciarse un desfase de estas últimas en relación a las extranjeras. La consecuencia de este aumento de la complejidad del análisis estadístico, que según diversos autores no se ha acompañado de un aumento paralelo de los conocimientos de estadística de los profesionales<sup>29-32</sup>, es la disminución de la capacidad de los lectores para comprender y evaluar los resultados de los estudios de investigación. A pesar de la insistencia sobre este fenómeno en los últimos años, el interés de los investigadores en estas materias todavía es escaso, lo que hace necesario dar una mayor importancia a la formación de los profesionales en metodología y estadística<sup>11,14</sup>. Hubiera sido interesante evaluar la adecuación del análisis estadístico para poder comprobar si este aumento de complejidad obedece a un mejor uso de la estadística o tan sólo refleja la mayor facilidad de utilización de programas informáticos sofisticados.

En un artículo científico, la presentación de los resultados debe reflejar los objetivos de la investigación. En este estudio se ha observado que el 51 % de los artículos presentan los resultados de forma adecuada. Probablemente este porcenta-

je hubiera sido menor si se hubieran evaluado otros aspectos de la presentación estadística de resultados, como la presentación del valor exacto de la p, o la existencia de una p huérfana. De hecho, Mora et al<sup>33</sup> al analizar estos aspectos encontraron incorrecciones mayores en el 15 % y menores en el 82 % de los artículos publicados en *Medicina Clínica* en 1993.

El uso de los intervalos de confianza ofrece grandes ventajas en relación con la significación estadística en las publicaciones biomédicas, ya que permite conocer la estimación de la magnitud del efecto observado. Por este motivo, diversos autores y comités editoriales de revistas han recomendado su uso en las publicaciones<sup>21</sup>. Sin embargo, el porcentaje de artículos revisados que utilizaban IC es tan sólo del 22,2 % en el global de la muestra y del 29,2 % en el año 1990. Estas cifras son similares a las observadas por Liberati et al<sup>34</sup>, e intermedias entre las encontradas por otros autores que oscilan desde el 3,6<sup>27</sup> y el 16,1 %<sup>26</sup> hasta el 51 %<sup>24</sup>. Recientemente, Mora et al<sup>33</sup> encontraron que el 43 % de los artículos publicados en *Medicina Clínica* en 1993 presentaban intervalos de confianza.

En este estudio se ha observado que en el 49,1 % de los artículos se observan diferencias estadísticamente significativas a favor de la terapia de estudio, porcentaje semejante al de Jiménez Villa et al<sup>35</sup>, mientras que en el 46,6 % no se aprecian diferencias estadísticamente significativas. Aunque este porcentaje puede ser considerado elevado, es similar al informado por otros autores<sup>36,37</sup>, y podría atribuirse tanto al insuficiente número de sujetos estudiados, como al incremento de estudios que comparan fármacos de la misma familia, que buscan demostrar su equivalencia en vez de su diferencia. Sin embargo, en cualquier caso, en un estudio negativo debe determinarse su potencia estadística para poder interpretarlo adecuadamente. En este estudio se ha encontrado que tan sólo el 9 % de los artículos realizan este cálculo, corroborando los datos de otros autores<sup>12,23,37</sup>.

En este estudio se ha puesto de manifiesto la insuficiente información sobre la metodología estadística en los artículos de investigación clínica de medicamentos realizados en centros españoles, especialmente cuando son publicados en revistas españolas. Aunque se aprecia una mejoría en los años evaluados en la mayoría de los ítems, la situación todavía dista mucho de ser la adecuada. Si se tiene en cuenta que la literatura biomédica es la principal vía de difusión de los avances terapéuticos, es preciso que tanto los investigadores como los comités editoriales de las

revistas adopten estrategias para mejorar no sólo la calidad de los estudios sino también de las publicaciones, proporcionando toda la información necesaria para que los lectores puedan evaluar la validez científica y la utilidad de los resultados. Así mismo, los profesionales sanitarios deberían tener la suficiente formación para realizar esta lectura crítica y evaluar la aplicabilidad de los resultados a los propios pacientes. Estudios como el presente pueden ayudar a prever a los departamentos de bioestadística de las facultades de medicina las necesidades de la formación continuada en estos temas para los profesionales de las ciencias de la salud.

## BIBLIOGRAFÍA

- Schor S, Karten I. Statistical evaluation of medical journals manuscripts. *JAMA* 1966; 195: 145-115.
- Hokanson JA, Luttmann DJ, Weiss GB. Frequency and diversity of use statistical techniques in *Oncology Journals*. *Cancer Treat Rep* 1986; 70: 589-594.
- Hokanson JA, Stiernberg CM, McCracken MS, Quinn FB. The reporting of statistical techniques in *Otorrinolaryngology Journals*. *Arch Otolaryngol Head Neck Surg* 1987; 113: 45-50.
- Hokanson JA, Ladoulis CT, Quinn FB, Bienkowski AC. Statistical techniques reported in *Pathology Journals* during 1983-1985: implications for *Pathology Educators*. *Arch Pathol Lab Med* 1987; 111: 202-207.
- Meinert CL, Tonascia S, Higgins K. Content of reports on *Clinical Trials*: A critical Review. *Control Clin Trials* 1984; 5: 328-347.
- Feinstein AR. Clinical bioestatistics. XXV. A survey of the statistical procedures in in general medical journals. *Clin Pharmacol Ther* 1974; 15: 97-107.
- Altman DG. Statistics in medical journals. *Stat Med* 1982; 1: 59-71.
- Murray GD. Statistical guidelines for the *British Journal of Surgery*. *Br J Surg* 1991; 78: 782-784.
- Nolasco A, Gascón E, Mur P, Ferrándiz E, Álvarez-Dardet C. Utilización de la estadística en publicaciones médicas: una comparación internacional. *Med Clin (Barc)* 1986; 86: 841-844.
- Farrús M, De la Fuente JA, Iglesias M, Borrell C. Revisión de los artículos originales publicados en *Atención Primaria* en Cataluña. *Aten Primaria* 1993; 12: 325-332.
- Martín M, Sanz F, Andreu D. Efecto de la introducción de la Bioestadística en el curriculum de los estudios de Medicina. Análisis de una década en la revista *Medicina Clínica (Barcelona)*. *Med Clin (Barc)* 1982; 79: 273-276.
- Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. *N Engl J Med* 1983; 309: 709-713.
- Pocock SJ, Hugges MD, Lee RJ. Statistical problems in reporting of clinical trials. *N Engl J Med* 1987; 317: 426-432.
- Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970; 3: 637-664.
- Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines or contributors to medical journals. *Br Med J* 1983; 286: 1.489-1.493.
- Altman DG. Statistics in Medical Journals: developments in the 1980's. *Stat Med* 1991; 10: 1.887-1.913.
- Álvarez-Dardet C, Gascón E, Mur P, Nolasco A. 10 years trends in the *Journal's* publications. *N Engl J Med* 1985; 312: 1.521-1.522.
- Bombardier C, Tugwell P, Sinclair A, Dok C, Anderso G, Buchanan WW. Preference for endpoint measures in clinical trials: results of structured workshops. *J Rheumatol* 1982; 9: 798-801.
- Smith DG, Clemens J, Crede W, Harvey M, Gracely EJ. Impact of multiple comparisons in randomized clinical trials. *Am J Med* 1987; 83: 545-550.
- Emrich LJ. Common problems with statistical aspects of periodontal research papers. *J Periodontol* 1990; 61: 206-208.
- Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 1986; 292: 746-750.
- González JC, Pulido M, Sanz F. Evaluación del uso de procedimientos estadísticos en los artículos originales publicados en *Medicina Clínica* durante tres décadas. *Med Clin (Barc)* 1995; 12: 448-452.
- Der Simonian R, Charette J, McPeck B, Mosteller F. Reporting methods in clinical trials. *N Engl J Med* 1982; 306: 1.332-1.337.
- Powe NR, Kinnison ML, Steinberg EP. Quality assessment of randomized controlled trials of contrast media. *Radiology* 1989; 170: 377-380.
- Tyson JE, Furzan JA, Reisch JS, Mize SG. An evaluation of the quality of therapeutic studies in perinatal medicine. *J Pediatrics* 1983; 102: 10-13.
- Jiménez Villa J, Carré Llopis MC, Argimón Pallás JM. Información sobre los aspectos metodológicos en los artículos de la revista *Atención Primaria*. *Aten Primaria* 1994; 14: 1.113-1.117.
- Evans M, Pollock AV. Trials on trials. A review of trials of antibiotic prophylaxis. *Arch Surg* 1984; 119: 109-113.
- Mora Ripoll R, Ascaso Terrén C, Sentís Vilalta J. Tendencias actuales en la utilización en la estadística en medicina. Estudios de artículos originales publicados en *Medicina Clínica (1991-1992)*. *Med Clin (Barc)* 1995; 105: 444-447.
- Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. *Am J Med Educ* 1980; 55: 692-697.
- Haynes SJ. Randomized clinical trials in neurosurgery. *Neurosurgery* 1983; 12: 259-264.
- Reznick RK, Dawson-Saunders E, Folse Jr. A rationale for the teaching of stistics to surgical residents. *Surgery* 1987; 101: 611-617.
- Pérez Mateo M, Boix V, Martínez A, Latour J. Análisis crítico de los ensayos clínicos en pancreatitis aguda publicados entre 1983 y 1989. *Med Clin (Barc)* 1991; 96: 371-373.

33. Mora Ripoll R, Ascaso Terren C, Sentís Vilalta J. Uso y presentación de la metodología estadística en los artículos originales publicados en Medicina Clínica durante 1993. *Med Clin (Barc)* 1995; 105: 9-12.
34. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4: 942-951.
35. Jiménez Villa J, Carré Llopis MC, Argimón JM. Tipos de estudios publicados en la revista Atención Primaria (1984-1990). *Aten Primaria* 1993; 12: 9-14.
36. Reed III F, Slaichert W. Statistical proof in inconclusive "negative" trials. *Ann Intern Med* 1981; 141: 1.307-1.310.
37. Friedman JA, Chalmers TC, Smith H, Kuebler RR. The importance of Beta, the type II error and sample size in the design and interpretation of the randomized control trial. Surveys of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-694.

---

# Diseño estadístico de ensayos clínicos

---

Carlos Rubio Terrés

Laboratorios Lederle. División de Cyanamid Ibérica, S.A. Madrid.

## El método en la investigación clínica

---

René Descartes (1596-1650) en su obra *Reglas para la dirección de la mente* estableció las directrices básicas del método científico. La regla IV indica que "para la investigación de la verdad de las cosas es necesario el método". La regla XIII propone que "si nosotros comprendemos perfectamente una cuestión, es preciso abstraerla de todo concepto superfluo, reducirla a su mayor simplicidad y dividirla en partes tan menudas como sea posible, enumerándolas"<sup>1</sup>. Estas reglas básicas, ideadas en el siglo XVII, siguen siendo válidas hoy día para el diseño de cualquier investigación y, por supuesto, para la investigación clínica.

La utilidad clínica de un tratamiento o medicamento se establece mediante dos tipos de razonamiento: el deductivo y el inductivo<sup>2</sup>. El razonamiento deductivo es el que va de lo general a lo específico. Es el que se plantea habitualmente el médico en la práctica clínica. Sabemos, por ejemplo, que un hipnótico determinado es un tratamiento eficaz del insomnio porque así ha sido demostrado en ensayos clínicos. Sin embargo, podemos preguntarnos, ¿cuál será su grado de eficacia en este paciente concreto? El razonamiento inductivo, por el contrario, va de lo específico a lo general. Supongamos que una compañía farmacéutica ha descubierto una molécula que tiene actividad hipnótico-sedante en animales y en voluntarios sanos. Pero, ¿es eficaz en el tratamiento del insomnio?, ¿es segura su utilización en dicha indicación?, ¿qué dosis son las adecuadas y en qué casos? La realización de estudios que respondan a éstas y a otras cuestiones similares es lo que se denomina investigación clínica. El método más utilizado es el ensayo clínico, es decir, aquel estudio experimental y prospectivo en el cual el investigador provoca y con-

trola las variables, y los pacientes son asignados de forma aleatoria a los distintos tratamientos que se comparan. Pero el ensayo clínico, entendido como el método científico así definido, es muy reciente. El primer ensayo clínico controlado y aleatorizado no se realizó hasta después de la Segunda Guerra Mundial, cuando Bradford Hill en 1946 estudió la eficacia de la estreptomina en la tuberculosis pulmonar<sup>3,4</sup>. En la historia de la medicina son abundantes los ejemplos de tratamientos establecidos sin el debido rigor científico o ético. Recientemente se ha sabido que en la década de 1940, en el Hospital de la Universidad de California (EE.UU.) se inyectó plutonio a numerosos pacientes supuestamente terminales, sin su conocimiento ni consentimiento, para obtener información acerca de sus efectos tóxicos, con fines militares<sup>5</sup>.

Otro conocido ejemplo es el del tratamiento de la esquizofrenia mediante el coma insulínico, propuesto por Manfred Sakel en 1933, sin suficiente base científica y desde luego sin realizar previamente estudios clínicos que demostraran la utilidad de dicho tratamiento<sup>6</sup>. A pesar de ello, fue de uso generalizado y aceptado hasta los años cincuenta, cuando la descripción de muertes y lesiones cerebrales en los pacientes tratados provocó la realización de un ensayo clínico cuyos resultados se publicaron en 1957. En el mismo no se demostró que el coma insulínico fuera superior al coma barbitúrico (que se supuso sin efecto) en la curación de la esquizofrenia<sup>7</sup>. Como consecuencia de dicho estudio, el tratamiento de la esquizofrenia mediante el coma insulínico fue progresivamente abandonado, con lo que se suprimió una práctica científica y probablemente dañina.

Queda clara, pues, la necesidad de realizar ensayos clínicos científica y éticamente válidos antes de instaurar un nuevo tratamiento. En la figura 1 se esquematiza la metodología general que debe seguirse en la realización de un ensayo clínico.

---

Publicado en *Med Clin (Barc)* 1996; 107 (8): 303-309.

## Definición del objetivo

Es el primer paso. Se trata simplemente de plantear una cuestión concreta, la hipótesis que se desea estudiar. Para ello debe realizarse una prueba de hipótesis. Si queremos comparar la efectividad de dos fármacos A y B, plantearemos una *hipótesis nula* ( $H_0$ ), es decir  $A \neq B$  (no habrá diferencias de eficacia entre ambos) y una *hipótesis alternativa* ( $H_1$ ):  $A = B$ , por tanto A será mejor o peor que B. De acuerdo con el resultado del estudio, aceptaremos o, por el contrario, rechazaremos la hipótesis nula<sup>8</sup>. Una hipótesis puede calificarse como “buena” cuando es “posible” en lo que se refiere al número de pacientes que deben seleccionarse, al tiempo de realización del estudio, y al coste y medios que se precisan para poder responderla. Así mismo, debe ser “relevante” porque se entiende que aportará nuevos conocimientos o redundará en mejoras de la práctica clínica, y debe tener suficiente “fundamento”, existiendo datos previos que la justifiquen. Por supuesto, debe ser de interés para el investigador y “ética” en su planteamiento<sup>9</sup>.

## Necesidad de grupo control

Es evidente que, en términos generales, para averiguar si un medicamento A es eficaz en una patología determinada, lo mejor es compararlo inicialmente con un tratamiento no activo (*placebo*). Una vez demostrado que A es mejor que placebo, para determinar en qué medida es eficaz, lo lógico es compararlo con un medicamento B *activo* y de eficacia ya conocida en esa patología. La incorporación de un grupo control tratado con placebo es importante también para establecer la comparación relativa de eficacia entre ambos tratamientos activos<sup>10</sup>.

La comparación frente a un grupo *no tratado* sólo está justificada cuando no es posible la administración de un placebo por motivos éticos (p. ej., su administración continuada por vía parenteral, o debido a la gravedad de la patología, como en una septicemia) o bien porque es posible la valoración objetiva de los resultados (p. ej., mediante hemocultivo).

Además del tratamiento activo, del placebo y del no tratamiento, pueden realizarse estudios controlados comparando con *otras dosis* (estudios de búsqueda de dosis) o bien con *controles históricos*<sup>10</sup>.

En estos últimos estudios surge la duda acerca de si la distribución de los factores pronósticos en el grupo histórico es similar o no a la del grupo contemporáneo. Para que los controles históricos

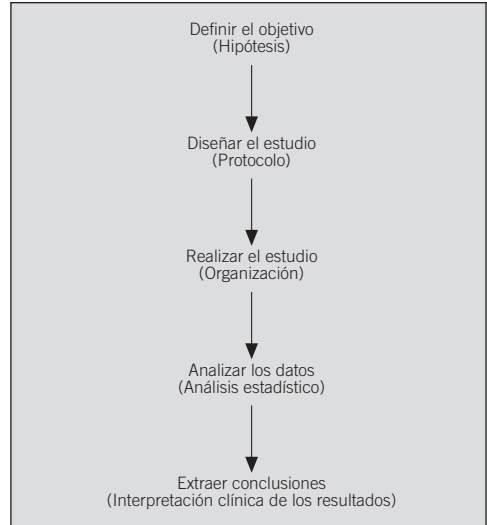


Fig. 1. El método en la investigación clínica.

puedan considerarse comparables a los controles contemporáneos, los pacientes deben proceder de los mismos hospitales y los métodos de evaluación de la respuesta, así como los criterios de inclusión/exclusión deben ser idénticos en ambos grupos. Otro aspecto importante que hay que tener en cuenta es que la enfermedad debe tener un desenlace previsible, sin posibilidad de mejoría o curación espontánea que podría diferir en el tiempo y llevar a conclusiones erróneas. No obstante, debe tenerse en cuenta que a pesar de la adopción de medidas como las citadas, los estudios con controles históricos tienden a exagerar el valor de un nuevo tratamiento<sup>11</sup>, por lo que sólo deben efectuarse por imposibilidad de establecer controles prospectivos, por ejemplo en enfermedades de muy baja incidencia, o cuando se prevean problemas serios en la selección de pacientes.

Los *autocontroles* (que cada paciente sea su propio control) pueden ser de utilidad para conocer la variabilidad intraindividual en determinados estudios, como en los de diseño cruzado de búsqueda de dosis. Sin embargo, para la evaluación de eficacia es preferible comparar distintos grupos de tratamiento.

Como principio, la realización de *estudios no controlados* debe ser la excepción. No obstante, éstos están justificados en determinados casos. No sería ético hacer estudios frente a placebo en pacientes con cáncer y privarles del tratamiento activo. No son controlados los estudios de farmacocinética en fase I. Tampoco algunos estu-



dios iniciales en fase II. Respecto a los estudios de seguridad en fase IV, es muy conveniente que sean controlados ya que el perfil de seguridad de un medicamento sólo puede establecerse en comparación con placebo (que también puede inducir efectos adversos) o con un fármaco de referencia<sup>12</sup>.

### Significación y precisión estadísticas. Relevancia clínica

Supongamos que una vez terminado el ensayo clínico en el que comparamos la eficacia de los medicamentos A y B, obtenemos una tasa de respuesta del 50 % con A y del 45 % con B. Realizada la correspondiente prueba estadística, resulta que A es mejor que B, con una  $p < 0,05$  y un IC 95 % = 2,5 al 9 %. ¿Qué es lo que esto significa? El principal objetivo de las pruebas estadísticas es responder a la pregunta ¿cuál es la probabilidad de que la diferencia observada se deba al azar?<sup>2,13</sup>. Una  $p < 0,05$  indica que en menos de cinco veces de cada 100 que repitiéramos el mismo estudio, nuestro resultado se debería al azar. Una  $p$  significativa indica que existe una gran diferencia de efectividad entre ambos fármacos, o bien que la muestra estudiada es tan grande que hemos conseguido detectar una diferencia real, pero pequeña<sup>13</sup>. Debe advertirse que la *significación estadística* a partir de  $p < 0,05$  es un convencionalismo, siendo también de interés la información que se obtiene de los resultados con una  $p$  algo mayor que 0,05. Cuanto mayor sea la  $p$ , más fuerte será la evidencia a favor de la hipótesis nula<sup>2</sup>.

A veces surge la duda de si debemos elegir una  $p$  con una o con 2 colas. La respuesta es simple: puede calcularse con una cola siempre y cuando consideremos que el medicamento A nunca podrá ser peor que el fármaco de referencia (p. ej., un placebo). Sin embargo, incluso en la comparación de un fármaco activo con placebo, cabe la posibilidad de que se produzca un resultado falso negativo (que no se hallen diferencias frente al placebo). Esto es más frecuente cuando la variable medida incluye criterios subjetivos, por ejemplo en la depresión psíquica, o cuando la muestra estudiada es pequeña<sup>10</sup>. Por tanto, generalmente es conveniente elegir una  $p$  con 2 colas (A puede ser mejor o peor que B). Esta decisión es importante, ya que una  $p = 0,04$  con una cola es estadísticamente significativa, mientras que la equivalente con 2 colas ( $p = 0,08$ ) no lo sería.

La diferencia encontrada entre dos tratamientos puede ser estadísticamente significativa, y sin

embargo no tener *relevancia clínica*. La relevancia clínica es un juicio subjetivo y pragmático acerca de la importancia real para la práctica clínica de la diferencia hallada. Por ejemplo, si comparásemos la tolerancia gastrointestinal de 2 antiinflamatorios A y B, el A no produjera trastornos gástricos en 90 de cada 100 pacientes y el B fuera, así mismo, bien tolerado en 93 de cada 100 tratados, aunque la diferencia fuese estadísticamente significativa ( $p < 0,05$ ) parece evidente que no tendría relevancia clínica<sup>13</sup>. La *precisión estadística* es la inversa de la variancia (a mayor precisión, menor variancia) que aumenta con el aumento del tamaño de la muestra. Una manera de medir la precisión es mediante el intervalo de confianza (IC). Cuanto más amplio es el IC menor es la precisión, hay menor confianza en que el resultado no se deba al azar y pueda inferirse a la población<sup>13</sup>.

### Errores en la investigación clínica

Un resultado estadísticamente significativo (la consabida  $p$ ) no supone ninguna garantía de que sea un resultado válido o real, que pueda inferirse a la población general con confianza, si no hemos cuidado el control de los errores estadísticos que se producen en los ensayos clínicos.

La *muestra* de pacientes incluidos en un ensayo clínico (p. ej., muestra de pacientes con infecciones ginecológicas de los hospitales de la provincia de Madrid) proviene de una *población de muestreo* (p. ej., totalidad de pacientes con infecciones ginecológicas de los hospitales de la provincia) y ésta, a su vez, proviene de la *población objetivo* (p. ej., pacientes españolas con infecciones ginecológicas)<sup>2</sup>. Se pretende que los resultados obtenidos en la muestra de pacientes puedan ser inferidos (generalizados) a la población objetivo. Para ello, deben controlarse los errores aleatorios y sistemáticos (o sesgos). Debe aclararse que ningún estudio está libre de errores y, por lo tanto, las inferencias nunca son perfectamente válidas<sup>9</sup>. Sin embargo, el objetivo es controlar al máximo posible dichos errores, para conseguir que:

1. La muestra seleccionada sea representativa de la población. Para ello, deben controlarse los *errores aleatorios*, aquellos que dan lugar a resultados erróneos debidos al azar. Su descontrol puede producir la distorsión de la muestra seleccionada (que la proporción de sus factores pronósticos sea diferente a la de la población). Por ejemplo, suponiendo que la prevalencia de sida en la población fuese del 30 %, en una muestra

bien seleccionada de individuos hallaríamos 30 con sida. Si en su lugar, el número de individuos con sida fuera de 20 o 40, esta distorsión podría deberse a un error aleatorio<sup>2,9</sup>.

Como consecuencia del control de los errores aleatorios, se produce el aumento de la *precisión* y de la *significación* estadísticas de una estimación. Pero, ¿cómo controlar este tipo de error?: mediante el *muestreo al azar* y *aumentando el tamaño de la muestra* ( $n$ )<sup>2,9</sup>.

2. La asignación de los pacientes de la muestra a los grupos de tratamiento debe dar lugar a una distribución similar de los factores pronósticos.

Deben controlarse, por tanto, los *sesgos* (*errores sistemáticos*), aquellos que producen resultados erróneos por defectos en el diseño del estudio. El grado de ausencia de sesgos va a determinar lo que se denomina *la validez estadística* del estudio. La *validez interna*, es decir, el grado de validez del resultado para los propios pacientes del estudio, y la *validez externa* o grado de confianza son la extrapolación o inferencia de ese resultado a la población<sup>14</sup>.

### Control de sesgos: aleatorización y enmascaramiento

Existen varios tipos de sesgos. Quizás uno de los principales es el de la *selección de pacientes* en el muestreo y en la asignación a los grupos de tratamiento<sup>14</sup>. Para controlar los sesgos del *muestreo* deben definirse correctamente los *criterios de inclusión/exclusión de pacientes*, para controlar adecuadamente los factores pronósticos conocidos (p. ej., en un ensayo clínico de un antimicrobiano en infecciones ginecológicas, serían posibles criterios de inclusión: mujeres hospitalizadas, de edad a partir de 16 años, con infecciones ginecológicas confirmadas mediante un diagnóstico clínico o bacteriológico y originadas por microorganismos patógenos de sensibilidad conocida o probable a el/los antimicrobianos en estudio; serían criterios de exclusión: pacientes alérgicas a antimicrobianos similares, disfunción renal de cierta intensidad, trastornos hepáticos, haber tomado otro antimicrobiano en las 72 h previas al reclutamiento, etc.). Así mismo, debe hacerse el muestreo *al azar*, con lo que se consigue el control de los factores pronósticos desconocidos (en el ejemplo anterior, supongamos que pueda existir un sector de la población femenina con una determinada característica o factor pronóstico desconocido que determine los resultados del estudio en infecciones ginecológicas; mediante el muestreo al azar, ese factor ten-

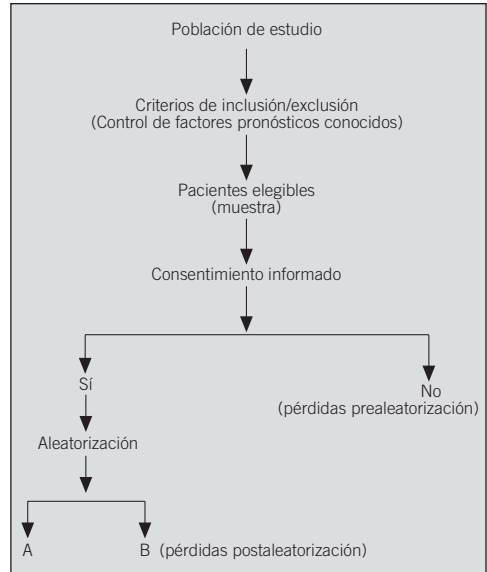


Fig. 2. Aleatorización: método general.

dría a estar en la muestra en igual proporción que en la población).

El otro sesgo en la selección es el que puede producirse en la asignación de los pacientes de la muestra a los grupos de tratamiento. Puede ocurrir que, de manera inconsciente o no, el investigador asigne a los pacientes con peor pronóstico al tratamiento ya conocido, en el que confía, y a los pacientes en mejor estado al tratamiento que se está probando. Para evitar este sesgo y, por tanto, con el objetivo de conseguir una distribución homogénea de los factores pronósticos, tanto de los conocidos como de los desconocidos, debe hacerse una *asignación aleatoria de los pacientes a los grupos de tratamiento* (*aleatorización*). El método general de la aleatorización se esquematiza en la figura 2. Con la asignación aleatoria no sólo se evitan los sesgos de asignación, sino que también se minimizan los propios errores aleatorios.

Los principales tipos y métodos de aleatorización (simple, por bloques, estratificada, centralizada, ciega y proporcional) son bien conocidos y están claramente descritos en la literatura<sup>15,16</sup>. Aunque debe valorarse caso por caso, en general puede decirse que la aleatorización por bloques es preferible a la simple, porque el número de pacientes por grupo es balanceado en la primera, y que, así mismo, es muy conveniente que sea ciega (la asignación de los tratamientos es descono-

cida para el investigador: estudios a doble ciego)<sup>15</sup>. Esto nos lleva al *sesgo de evaluación de los resultados* y a su control mediante el *enmascaramiento*. En un ensayo clínico no ciego (nadie –médico o paciente– desconoce la asignación de los tratamientos en estudio) es fácil que se introduzcan distorsiones (sesgos) en los efectos y valoraciones condicionadas –subjetivas y falsas– del tratamiento y de los resultados, tanto por el paciente como por parte del médico. Para evitarlo, se recurre al enmascaramiento, realizándose los denominados estudios a simple ciego (tratamiento desconocido por el paciente), a doble ciego (también desconocido para el médico), a triple ciego (también para el monitor) y cuádruple ciego (también en el análisis estadístico)<sup>15</sup>.

Los estudios no enmascarados pueden estar justificados en casos muy concretos: por imposibilidad galénica de realizar el enmascaramiento, por motivos éticos (p. ej., no es posible la administración parenteral continuada de un placebo), por las características del fármaco (p. ej., estudio de un antineoplásico del que deben hacerse ajustes de dosis).

### **El impacto de las pérdidas pre y postaleatorización**

Como se mencionó anteriormente, los objetivos de la asignación aleatoria de los pacientes a los grupos de tratamiento son fundamentalmente dos: evitar los sesgos de asignación y minimizar los errores aleatorios.

Charlson et al<sup>17</sup> revisaron 41 ensayos clínicos aleatorizados existentes en la base de datos del Instituto Nacional de la Salud de EE.UU. hasta 1979, hallando que las elevadas *pérdidas prealeatorización* (por la negativa de los pacientes a entrar en el estudio, o bien por ser considerados no elegibles de acuerdo con los criterios de inclusión/exclusión) (fig. 2) fueron responsables de que tan sólo el 34 % de los estudios alcanzara la *n* prevista inicialmente. Estas pérdidas, además, disminuyen la validez externa de los estudios, ya que surge la duda acerca de si los pacientes no aleatorizados de la población de muestreo pueden ser diferentes en algún factor pronóstico respecto a los aleatorizados. Por ello, es muy importante hacer el seguimiento de los pacientes elegibles no aleatorizados<sup>18</sup>.

¿Pero, de qué manera pueden reducirse estas pérdidas de pacientes antes de la aleatorización? En primer lugar, cuando sea posible, reduciendo el número de pacientes no elegibles, mediante criterios de inclusión y exclusión menos estrictos, es decir, realizando estudios más pragmáticos.

En segundo lugar, consiguiendo que sea aleatorizado el mayor tanto por ciento posible de los pacientes elegibles (que cumplen los criterios de selección). Esto es fácil de decir, pero difícil en la práctica, ya que se trata de plantear al paciente la solicitud de su consentimiento de una manera “inteligente”, para que, sin alarmarle pero al mismo tiempo informándole adecuadamente, se preste a participar en el ensayo clínico<sup>17</sup>.

Las *pérdidas postaleatorización* (fig. 2) debidas a los abandonos y retiradas de pacientes durante el estudio son, si cabe, más graves. Destruyen la aleatorización y, por tanto, la validez interna y externa del ensayo. El tratamiento de estas pérdidas debe estar claramente especificado en el protocolo. Así mismo, deben tenerse en cuenta en el cálculo del tamaño de la muestra. Un método que evita la destrucción de la asignación aleatoria es el *análisis de los resultados por intención de tratar*, que incluye a todos los pacientes aleatorizados, hayan sido o no tratados. Este tipo de análisis es de gran importancia, ya que los resultados de los pacientes retirados pueden suministrar información muy valiosa. Un ejemplo es el del Helsinki Heart Study, acerca de la prevención primaria de la mortalidad coronaria mediante el hipolipemiante gemfibrozilo<sup>19</sup>. En el grupo tratado con el fármaco activo, se describió una tasa de mortalidad por accidentes, homicidios y suicidios algo superior a la observada en el grupo placebo<sup>20</sup>, lo que ha llevado a plantear hipótesis acerca de los posibles efectos psiquiátricos derivados de las concentraciones bajas de colesterol.

La visión explicativa de los resultados es la que tiende a hacerse habitualmente, y consiste en el análisis sólo de los pacientes evaluables (no contempla los abandonos y las retiradas del estudio). La visión pragmática, como se ha dicho, es la del análisis por intención de tratar, de todos los pacientes aleatorizados. Ambos análisis (explicativo y pragmático) no son excluyentes y, por el contrario, puede ser de interés compararlos (tabla I). La visión superpragmática consistiría en analizar a todos los pacientes elegibles (aleatorizados y no aleatorizados). Sin embargo, este tipo de análisis parece utópico, paradójicamente desde un punto de vista práctico, por lo que generalmente nos conformaremos con el seguimiento de las pérdidas prealeatorización, tal y como se explicó con anterioridad.

### **El desequilibrio de factores pronósticos**

A pesar de la aleatorización, por errores en la misma o bien debido al propio azar, puede ocurrir que al analizar los factores pronósticos (p. ej.,

TABLA I  
ENSAYOS CLÍNICOS EXPLICATIVOS Y PRAGMÁTICOS<sup>a</sup>

<i>Característica</i>	<i>Ensayo explicativo</i>	<i>Ensayo pragmático</i>
Variable medida	Efecto del tratamiento en "condiciones de laboratorio"	Beneficio del tratamiento en la práctica clínica
Criterios de selección	Estrictos (pacientes homogéneos)	Laxos (todos los pacientes susceptibles de tratamiento)
Aleatorización	Sí	Sí
Enmascaramiento	Sí	No
Criterios de evaluación/objetivo	A > B ( $p < 0,05$ )	A es tratamiento de elección
Tipos de error	$\alpha$ y $\beta$	$\gamma$
Pacientes necesarios (n)	Para controlar $\alpha$ y $\beta$	Para controlar $\gamma$

<sup>a</sup>Tomados de referencia 31 (modificado).

edad, sexo, gravedad de la enfermedad, etc.) de los grupos de tratamiento, encontremos que existe un desequilibrio en alguno de ellos (p. ej., que haya más pacientes graves en un grupo que en otro, siendo la diferencia estadísticamente significativa). Este riesgo puede reducirse mediante la aleatorización estratificada. De ese modo, se asigna a los pacientes a unos subgrupos o estratos previamente establecidos, según factores pronósticos que puedan determinar los resultados del estudio<sup>15</sup>. Si en un estudio se considera que la edad es un factor pronóstico importante, puede estratificarse en pacientes de edad menor de 50 y de mayor o igual a 50 años, por ejemplo. En algún estudio puede ser de interés estratificar en pacientes ambulatorios y no ambulatorios<sup>11</sup>, pacientes que han sufrido sólo un infarto o más de uno, etc. Qué factor, si es que hay alguno, tiene suficiente relevancia como para ser estratificado debe decidirse en cada ensayo clínico.

Es importante estratificar sólo un número reducido de factores (generalmente dos o tres como máximo) ya que la introducción de un número excesivo de variables a estratificar aumenta la complejidad del estudio y puede dificultar la consecución del número necesario de pacientes en cada estrato<sup>15</sup>.

La evaluación del posible desequilibrio en los factores pronósticos es fundamental, lógicamente, en los estudios no aleatorizados, ya que en éstos el riesgo de desequilibrio es muy elevado.

Si se confirma el desequilibrio, debe realizarse el *ajuste de los factores pronósticos*, para comprobar si la diferencia entre los tratamientos sigue siendo significativa después del ajuste. El método de ajuste más utilizado para desequilibrios en

uno o dos factores es el test de Mantel-Haenszel. En el caso de observarse diferencias significativas en tres o más de tres factores, para variables cuantitativas se aplica la regresión múltiple y para variables cualitativas la regresión logística<sup>11</sup>.

Cuando no se hallan diferencias entre los resultados, analizados por intención de tratar, puede ser de interés realizar *análisis de subgrupos*. Estos análisis no son, en ningún caso, sustitutorios de la aleatorización estratificada y deben realizarse con prudencia, ya que las diferencias halladas en los subgrupos pueden llevar a conclusiones erróneas<sup>12</sup>. Un ejemplo es el del estudio The Multiple Risk Factor Intervention Trial (MRFIT), que no demostró un beneficio global originado por la intervención y, sin embargo, en el análisis de subgrupos se observó reducción de la mortalidad cardiovascular al cabo de un año en relación con la disminución del colesterol y el tabaquismo y, así mismo, en relación con la disminución de la tensión arterial. Estos resultados deben tomarse con gran prudencia<sup>21,22</sup>.

### El efecto del tamaño muestral

Una vez finalizado un estudio (A frente a B) podemos obtener tres resultados posibles: A es igual a B, o bien A es distinto de B (A es mejor o peor que B). Pero cualquiera de los resultados pudiera ser falso, es decir: el resultado del estudio puede no corresponder a la realidad. Pueden, por tanto, producirse resultados *falsos positivos* y *falsos negativos* (tabla II).

Si encontramos una diferencia (p. ej.,  $p < 0,05$ , con un IC del 95 %) el *nivel de confianza* ( $1-\alpha$ ) será la probabilidad de que el IC contenga el valor real. La probabilidad de que la diferencia ha-

TABLA II  
RIESGOS DE RESULTADOS FALSOS  
EN UN ENSAYO CLÍNICO

Resultado en el ensayo clínico	Realidad		
	A > B	A = B	A < B
A > B	-	1/2 $\alpha$	$\gamma$
A = B	$\beta$	-	$\beta$
A < B	$\gamma$	1/2 $\alpha$	-

A > B: A mejor que B; A = B: A igual a B; A < B: A peor que B;  $\alpha$ : falso positivo (riesgo  $\alpha$ /error de tipo I);  $\beta$ : falso negativo (riesgo  $\beta$ /error de tipo II);  $\gamma$ : falso positivo pragmático (riesgo  $\gamma$ /error de tipo III).

llada y contenida en el IC no sea real es lo que se denomina *error de tipo I* o riesgo de *falso positivo*<sup>23</sup>  $\alpha$ . El riesgo  $\alpha$  se reduce y, por tanto, aumentan el nivel de confianza  $1-\alpha$  y la precisión del IC mediante el aumento del tamaño muestral (n)<sup>23</sup>. Consideración especial merecen los *ensayos clínicos negativos* ( $p > 0,05$ ). Un resultado negativo (no se hallan diferencias) puede deberse a que realmente no exista diferencia entre los tratamientos o a que ésta sea nimia y prácticamente indetectable. En ese caso, se trata de un verdadero negativo. Si, por el contrario, es un resultado *falso negativo* (no se detecta una diferencia que existe) puede deberse a dos motivos<sup>23</sup>: a) la existencia de sesgos en el estudio (en la selección de los pacientes, en la información recogida o en su análisis), o bien b) la falta de precisión estadística, debido a la existencia de un *error aleatorio de tipo II*, cuya probabilidad se denomina  $\beta$  (tabla II). En este caso, se dice que el estudio carece del suficiente *poder estadístico* ( $1-\beta$ ). El poder estadístico aumenta con el aumento de n.

¿Pero qué hacer ante un ensayo clínico negativo? En el ejemplo<sup>24</sup> de la tabla III, observamos que se efectuó un análisis intermedio con tan

TABLA III  
TASA DE RESPUESTA A UN  
TRATAMIENTO: INTERVALO DE  
CONFIANZA Y PODER ESTADÍSTICO<sup>24</sup>

Tratamiento	Análisis intermedio	Análisis final
A	13/25 (52 %)	500/1.000 (50 %)
B	8/25 (32 %)	450/1.000 (45 %)
p	0,25	0,03
IC 95 %	-7 al 47 %	0,6 al 9 %

IC 95 %: intervalo de confianza del 95 %.

sólo 25 pacientes tratados en cada grupo. La diferencia hallada no fue estadísticamente significativa y el IC 95 % incluyó al 0, lo que confirma que no existía en ese momento una diferencia real entre A y B. Sin embargo, cuando se alcanzó la muestra deseada ( $n = 1.000$  pacientes por grupo), entonces la diferencia sí fue significativa ( $p = 0,03$ ), aunque el IC 95 % estuvo cerca del 0 (0,6-9 %) por lo que persisten algunas dudas acerca de la realidad de la diferencia hallada<sup>24</sup>. Vemos, así, la importancia de valorar la precisión (el intervalo de confianza) a la hora de analizar un resultado negativo (o positivo) y, así mismo, el efecto del tamaño muestral. Por tanto, ante un ensayo clínico negativo, las 2 preguntas que debemos hacernos son<sup>23</sup>: a) ¿qué información nos proporciona el intervalo de confianza?, y b) ¿es el número de pacientes estudiados suficiente para encontrar diferencias reales? En este caso, debe calcularse el poder estadístico *a posteriori* ( $1-\beta$ ). Se considera que el poder del estudio es aceptable cuando está entre el 80 y el 90 %.

En algunas ocasiones se plantea la realización de *estudios de equivalencia terapéutica*, es decir: ensayos clínicos que buscan demostrar igualdad entre los tratamientos. En primer lugar, debe decirse que intentar “probar la hipótesis nula ( $H_0$ )” es estadísticamente erróneo<sup>23</sup>. La  $H_0$  que se plantea es, por ejemplo, que  $A \geq B + d$ , siendo A y B el tanto por ciento de éxitos en uno y otro grupo de tratamiento y “d” la diferencia máxima aceptable para confirmar la equivalencia. Si se rechaza la  $H_0$ , se considera que  $A = B$  porque no difieren más del valor “d”. El objetivo de este tipo de estudios es conseguir demostrar equivalencia con una n pequeña, pero esto es, así mismo, erróneo, ya que la n necesaria para demostrar equivalencia es mayor que la precisa para demostrar diferencias (cuanto mayor es la diferencia entre dos tratamientos, menor es el número de pacientes que se necesita para poder demostrarlo)<sup>23</sup>.

Puede ocurrir que en un análisis intermedio de un ensayo clínico, se observe que los tantos por ciento de respuesta hallados difieren de los inicialmente previstos y extraídos de la revisión de la bibliografía. En ese caso, podría estar justificado modificar el protocolo y prolongar el ensayo clínico, para alcanzar una muestra suficiente, capaz de detectar diferencias.

El *error de tipo III* (riesgo  $\gamma$ ) podría denominarse como *falso positivo “pragmático”*, ya que es el que se produce cuando el resultado del ensayo indica que, por ejemplo, A es mejor que B, cuando la realidad es todo lo contrario: que A es peor que B (tabla II). Esto es de suma impor-

tancia en un estudio pragmático que pretenda decidir cuál debe ser el tratamiento de elección (A o B). También se reduce el riesgo  $\gamma$  mediante el aumento del tamaño de la muestra ( $n$ ) (tabla II).

Hay numerosa bibliografía que se ocupa de revisar, con ejemplos, los métodos y fórmulas utilizados para el cálculo del tamaño muestral<sup>3,11,25,26</sup>. Aquí tan sólo comentaremos la fórmula general para el cálculo de  $n$  en la comparación de medias y porcentajes<sup>3,25</sup> y teniendo en cuenta los riesgos  $\alpha$  y  $\beta$  (no el riesgo  $\gamma$ ) válida, por lo tanto, para estudios explicativos (no para estudios pragmáticos) (v. tabla I, con las características de ambos tipos de estudios):

$$n = (2 \times V / \delta) \times f(\alpha, \beta)$$

donde  $n$  es el número total de individuos a estudiar en el ensayo.  $V$  es la variabilidad; si se trata de la diferencia de medias corresponde a  $(S_1^2 + S_2^2)$ ; si se trata de la diferencia de porcentajes equivale a  $(P_1 [100-P_1] + P_2 [100-P_2])$ .  $\delta$  es la diferencia (o nivel de sensibilidad) que se considera clínicamente relevante (p. ej.,  $P_1-P_2 =$  una reducción del 5 % en la mortalidad); este valor se evalúa y selecciona mediante la revisión de la bibliografía existente antes del inicio del estudio.  $f(\alpha, \beta)$  es una constante que depende del nivel de protección frente a los errores de tipo I y II (se obtiene mediante tablas).

### **Análisis intermedios: ¿cuándo parar un EC?**

La razón de ser de los análisis intermedios es que si, por ejemplo, en el estudio de dos tratamientos en una enfermedad grave (que produce mortalidad), se observan diferencias relevantes a favor de uno de ellos, por motivos éticos puede evitarse la prolongación innecesaria del estudio, consiguiendo que el grupo que recibe el tratamiento que parece menos favorable no se vea perjudicado. No obstante, este razonamiento es muy difícil de aplicar en la práctica<sup>27</sup>. Debe establecerse una, así llamada, *regla de interrupción*, consistente en fijar, en el protocolo del estudio, la variable que se analizará (p. ej., mortalidad), así como el número y momento de realización de análisis estadísticos intermedios que se llevarán a cabo a lo largo del estudio. Así mismo, se determinará el nivel de significación (o sensibilidad) que deberá alcanzar la diferencia entre los tratamientos, para efectuar la interrupción del estudio. Por ejemplo, de acuerdo con la regla de O'Brien<sup>27</sup> para mortalidad, deben realizarse 5 análisis en el plazo de 1 año, siendo los niveles

de sensibilidad para decretarse la interrupción, los siguientes:

- 1.º:  $p < 0,00000001$
- 2.º:  $p < 0,0001$
- 3.º:  $p < 0,001$
- 4.º:  $p < 0,004$
- 5.º:  $p < 0,009$

Existen otras reglas de interrupción, como la de Peto, consistente en 3 análisis<sup>27</sup>, todos con una  $p < 0,001$ .

Un aspecto importante que afecta a las reglas de interrupción es el de la *corrección del exceso de pruebas estadísticas*. Aunque en la realidad A sea igual que B, el aumento en el número de pruebas estadísticas aumenta el riesgo de error  $\alpha$  (de resultados falsos positivos)<sup>11</sup>. Por ejemplo, si repitiéramos 10 pruebas de significación al nivel del 5 %, el nivel acumulado de significación, que sería de 0,05 en la primera prueba, al cabo de las 10 pruebas ascendería a 0,19 (fácilmente alcanzable, sin que existan diferencias reales)<sup>11</sup>. Por ello, si quisiéramos una  $p$  acumulada = 0,05 y estuviera previsto hacer 10 análisis intermedios, cada uno debería tener una regla de interrupción de  $p < 0,01$ <sup>11</sup>.

En otro orden de cosas, debe comentarse brevemente la importancia de guardar la confidencialidad de los análisis intermedios. Deben ser ciegos para el investigador, con el propósito de evitar futuros sesgos en el caso de que el ensayo clínico prosiga. El investigador sólo debe conocer el resultado del análisis si se decide interrumpir el estudio. La mejor manera de guardar dicha confidencialidad es mediante la creación de un Comité independiente que se encargue del análisis. Así mismo, debe evitarse la publicación de los resultados mientras esté en marcha la aleatorización.

Un ejemplo reciente de la problemática de los análisis intermedios es el de la zidovudina en el tratamiento del sida<sup>28</sup>. Los primeros ensayos clínicos aleatorizados que se realizaron con este fármaco fueron interrumpidos porque se hallaron diferencias significativas frente al placebo en el recuento de CD4 por debajo de 500/ $\mu$ l. En 1989, de acuerdo con estos resultados y debido a la presión de los grupos activistas del sida en EE.UU., la FDA decidió la aprobación de la zidovudina en los infectados por VIH asintomáticos<sup>28</sup>. Sin embargo, los resultados preliminares publicados con posterioridad, en 1993, correspondientes a un estudio europeo que no se interrumpió sino que prosiguió, el European Concorde Trial<sup>29</sup>, no mostraron una diferencia

clara a favor de la zidovudina en comparación con el placebo después de 3 años de seguimiento, ni en la supervivencia ni en la progresión de la enfermedad<sup>29</sup>. Este resultado hizo que en 1993 los activistas del sida de EE.UU. pidieran a la FDA que, en lo sucesivo, retrasase la aprobación de nuevos agentes anti-VIH<sup>28</sup>. Este caso ilustra claramente el conflicto existente entre la responsabilidad de aleatorizar a un paciente individual a un tratamiento que parece ser peor inicialmente, y la responsabilidad o el compromiso de asegurarse del resultado, prosiguiendo el estudio, lo que redundará en el beneficio de miles o millones de futuros pacientes.

## Conclusiones

La metodología estadística de la investigación clínica es motivo de preocupación y muestra de ello es la reciente publicación de un borrador de Guía al respecto del Comité de Especialidades Farmacéuticas (CEF o CPMP) de la Comisión de la Unión Europea<sup>30</sup>. Al comienzo de esta revisión, hablábamos del método científico y de su importancia en la investigación clínica. Es cierto que la metodología clínica y estadística y las buenas prácticas clínicas han complicado sobremedida la realización de ensayos clínicos. Sin embargo, debe tenerse siempre presente que el sujeto de la investigación es el ser humano y que, por tanto, los presupuestos éticos deben ir unidos a los metodológicos. Un ensayo clínico mal diseñado nunca debe hacerse ya que en ese caso se estaría incurriendo en el tan denostado "cobayismo". No obstante, la complejidad nunca deberá alcanzar cotas tales que impidan (por su elevado coste) el desarrollo de nuevos fármacos de interés sanitario. Sería, tal vez, útil profundizar en el papel de los denominados estudios pragmáticos<sup>31</sup> (tabla I). En cualquier caso, y parafraseando a Séneca, deberemos convenir que "todo lo honesto es difícil"<sup>32</sup>, y la investigación clínica no es una excepción.

## BIBLIOGRAFÍA

1. Descartes R. Discurso del método. Reglas para la dirección de la mente. Barcelona: Ediciones Orbis, S.A., 1983.
2. Plasencia A, Porta Serra M. La calidad de la información clínica (II): significación estadística. *Med Clin (Barc)* 1988; 90: 122-126.
3. Bakke OM, Carné X, García Alonso F. Ensayos clínicos con medicamentos. Fundamentos básicos, metodología y práctica. Barcelona: Doyma, S.A., 1994.
4. Streptomycin in Tuberculosis Trial Committee. Streptomycin treatment of pulmonary tuberculosis. *Br Med J* 1948; 2: 769-782.
5. Herken G, David J. Doctors in ardent pursuit of radiation weapons. *International Herald Tribune* 1994; 14 de enero: 6.
6. Spector R. Introduction to therapeutics. En: Spector R, editor. *The scientific basis of clinical pharmacology. Principles and examples*. Boston: Little, Brown and Co., 1986; 3-11.
7. Ackner B, Harris A, Oldham AJ. Insulin treatment of schizophrenia: a controlled trial. *Lancet* 1957; 1: 607.
8. González AG. Diseño y cálculo de tests estadísticos para ensayos clínicos y de laboratorio. Madrid: Escuela Universitaria de Enfermería, Universidad Complutense de Madrid, 1989.
9. Cummings SR, Browner WS, Hulley SB. Conceiving the research question. En: Hulley SB, Cummings SR, editores. *Designing clinical research*. Baltimore: Williams & Wilkins, 1988; 12-17.
10. Makuch RW, Johnson MF. Dilemmas in the use of active control groups in clinical research. *IRB* 1989; 11: 1-5.
11. Pocock SJ. *Clinical trials. A practical approach*. Chichester: John Wiley & Sons, 1983.
12. Circular 18/90. Directrices para la realización de estudios de Farmacovigilancia. Madrid: Dirección General de Farmacia y Productos Sanitarios, 21 de noviembre, 1990.
13. Porta Serra M, Plasencia A, Sanz F. La calidad de la información clínica (y III): ¿estadísticamente significativo o clínicamente importante? *Med Clin (Barc)* 1988; 90: 436-468.
14. Porta Serra M, Álvarez-Dardet C, Bolúmar F, Plasencia A, Veilla E. La calidad de la información clínica (I): validez. *Med Clin (Barc)* 1987; 89: 741-747.
15. Galende I, Tristán C. Problemas prácticos en un ensayo clínico (II). En: García Alonso F, Bakke OM, editores. *Metodología del ensayo clínico*. Monografía Dr. Antonio Esteve, n° 11. Barcelona: Fundación Dr. Antonio Esteve, 1991; 21-30.
16. Zelen M. The randomization and stratification of patients to clinical trials. *J Chron Dis* 1974; 27: 365-375.
17. Charlson ME, Horwitz R. Applying results of randomized trials to clinical practice: impact of losses before randomisation. *Br Med J* 1984; 289: 1.281-1.289.
18. De Abajo FJ, Serrano Castro MA. Problemas prácticos en un ensayo clínico (I). En: García Alonso F, Bakke OM, editores. *Metodología del ensayo clínico*. Monografía Dr. Antonio Esteve, n° 11. Barcelona: Fundación Dr. Antonio Esteve, 1991; 11-20.
19. Frick MH, Elo O, Haapa K, Heinonen OP, Heinsalmi P, Helo P et al. Helsinki Heart Study: primary-prevention trial with Gemfibrozil in middle-aged men with dyslipidemia. *N Engl J Med* 1987; 317: 1.237-1.245.
20. Frick MH, Heinonen OP, Huttunen JK, Manninen V. Gemfibrozil and coronary heart disease. *N Engl J Med* 1988; 318: 1.257.

21. Bulpitt CJ. Subgroup analysis. *Lancet* 1988; 2: 31-34.
22. Three analysis issues pertinent to designing an experiment. En: Hulley SB, Cummings SR, editores. *Designing clinical research*. Baltimore: Williams & Wilkins, 1988; 212-214.
23. Porta Serra M, Moreno V, Sanz F, Carné X, Velilla E. Una cuestión de poder. *Med Clin (Barc)* 1989; 92: 223-228.
24. Simon R. Confidence intervals for reporting results of clinical trials. *Ann Intern Med* 1986; 105: 429-435.
25. Carné X, Moreno V, Porta Serra M, Velilla E. El cálculo del número de pacientes necesario en la planificación de un estudio clínico. *Med Clin (Barc)* 1989; 92: 72-77.
26. Machin D, Campbell MJ. *Statistical tables for the design of clinical trials*. Oxford: Blackwell Scientific Publications, 1987.
27. Pocock SJ. When to stop a clinical trial. *Br Med J* 1992; 305: 235-240.
28. On stopping a trial before its time [editorial]. *Lancet* 1993; 342: 1.311-1.312.
29. Aboulker JP, Swart AM. Preliminary analysis of the Concorde Trial. *Lancet* 1993; 341: 889-890.
30. CPMP Working Party on Efficacy of medicinal products. Note for guidance: Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products (III/3630/92-EN, Draft 4). Bruselas: Commission of the European Communities, marzo 1993.
31. Vallvé C. Ensayos clínicos abiertos en investigación clínica. En: García Alonso F, Bakke OM, editores. *Metodología del ensayo clínico*. Monografía Dr. Antonio Esteve, n° 11. Barcelona: Fundación Dr. Antonio Esteve, 1991; 49-55.
32. Séneca LA. De los beneficios. En: Lucio Anneo Séneca: *Obras completas*. Madrid: Aguilar, 1961.



# La regresión a la media en la investigación y práctica clínica

Eliseo Guallar<sup>a</sup>, F. Javier Jiménez<sup>a</sup>, Fernando García-Alonso<sup>b</sup>  
y Olav M. Bakke<sup>c</sup>

<sup>a</sup>Departamento de Epidemiología y Bioestadística. Escuela Nacional de Sanidad. Instituto de Salud Carlos III. Madrid.

<sup>b</sup>Centro Nacional de Farmacobiología. Instituto de Salud Carlos III. Madrid. <sup>c</sup>Departamento de Farmacología. Institut Municipal d'Investigació Mèdica. Barcelona.

Al estudiar la importancia de los factores hereditarios en el crecimiento de las plantas del guisante, Sir Francis Galton observó hace más de 100 años que al cruzar las plantas de mayor tamaño entre sí se obtenían plantas por término medio más pequeñas, mientras que al cruzar las más pequeñas se originaban plantas relativamente mayores. Galton también observó este fenómeno en sus estudios sobre la herencia de la altura en humanos, denominándolo regresión a la mediocridad<sup>1</sup>. El fenómeno descrito por Galton es un caso particular de lo que en la actualidad se denomina regresión a la media<sup>2</sup>.

En la práctica, la regresión a la media se manifiesta como la tendencia de los valores extremos de una variable con el resultado de, por término medio, menos extremos en mediciones sucesivas<sup>3,4</sup>. Para que se produzca regresión a la media deben concurrir dos circunstancias: en primer lugar, debe producirse una selección de los valores de una variable (p. ej., seleccionar a los pacientes con cifras de colesterol sérico mayores de 6 mmol/l). En segundo lugar, debe existir un cierto grado de variabilidad intraindividual en las medidas sucesivas de la variable, bien como variabilidad fisiológica verdadera, bien como consecuencia de cambios en las circunstancias ambientales o de errores de medida aleatorios. Cuanto más extremo es el punto de corte y cuanto mayor es la variabilidad de la medida, mayor es el efecto de la regresión a la media<sup>3,5,6</sup>.

Aun a riesgo de simplificar en exceso el fenómeno, la regresión a la media puede entenderse mejor si se considera una variable para la que no exista variabilidad fisiológica intraindividual, pero sí errores de medida aleatorios. En estas circunstancias, al seleccionar individuos con valores extremos se incluye un grupo de pacientes cuyo valor real está por debajo del punto de cor-

te, pero que debido a errores de medida presentan valores superiores. Esto da lugar a errores en la clasificación de algunos sujetos así como a una sobreestimación del valor medio del grupo seleccionado<sup>7</sup>. Al repetir las medidas, se producirá una estimación más correcta del valor medio del grupo seleccionado (más cercano a la media de la población global)<sup>5,7</sup>. En la figura 1 se ilustra un ejemplo del efecto de la regresión a la media.

Aunque en la investigación clínica y epidemiológica este fenómeno ha sido estudiado desde hace tiempo<sup>5,7,8</sup>, la regresión a la media continúa siendo una fuente de error frecuentemente ignorada tanto en el diseño de estudios como en la interpretación de sus resultados. El objetivo de este trabajo es revisar de una manera intuitiva, y a partir de ejemplos extraídos de la literatura médica, los principales efectos de la regresión a la media, que se traducen en: a) inconsistencias en el diagnóstico y clasificación de pacientes; b) eva-

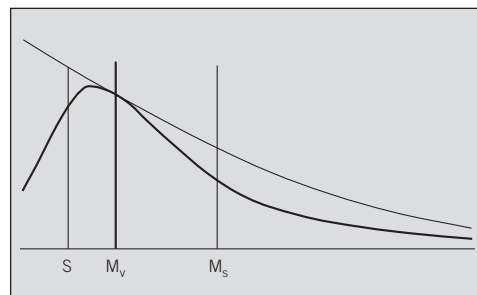


Fig. 1. Efecto de la regresión a la media en estudios clínicos y epidemiológicos. En trazo fino (—) se ilustra el extremo de una población seleccionada según un criterio de selección (S), así como la media de este grupo (M<sub>s</sub>). En trazo grueso (—) se ilustran los valores reales de estos sujetos, así como su media "verdadera" (M<sub>v</sub>), que se obtendría si se tomaran varias medidas de cada sujeto. La diferencia entre M<sub>s</sub> y M<sub>v</sub> es el efecto de la regresión a la media.

luación sesgada de la eficacia de las intervenciones clínicas; c) solapamiento con el efecto placebo; d) ineficiencia en la planificación de ensayos clínicos y estudios epidemiológicos, y e) errores de medida no diferenciales. Finalmente, discutiremos cómo minimizar los efectos de la regresión a la media y sus consecuencias en la investigación y la práctica médica.

### **Inconsistencia en el diagnóstico y clasificación de pacientes**

En la experiencia clínica es frecuente comprobar cómo algunos “enfermos” dejan de cumplir los criterios de enfermedad o pasan a ocupar otra categoría diagnóstica sin intervención externa aparente, dando lugar a inconsistencias en el diagnóstico o en la clasificación de los pacientes entre el momento de la selección y el seguimiento posterior. Muchas de estas inconsistencias son debidas a la regresión a la media y, más concretamente, a los errores de clasificación asociados. En este sentido, son múltiples las situaciones en las que se determina el estado de enfermedad de los pacientes en función de los resultados de una prueba diagnóstica. La presencia de errores de medida aleatorios y/o variabilidad de los parámetros fisiológicos da lugar a la existencia de individuos que, sin estar enfermos, presentan en el momento de la cuantificación cifras superiores a las utilizadas como criterio diagnóstico y que, por tanto, van a ser erróneamente clasificados como enfermos. En el seguimiento posterior, los casos que presentaban cifras extremas tienden a encontrarse más cerca de sus valores habituales, dejando de cumplir los criterios diagnósticos o de clasificación.

Como ejemplo ilustrativo, la regresión a la media puede explicar parte de la denominada “hipertensión de bata blanca”<sup>9</sup>. Este síndrome es el resultado de una hiperreactividad presora producida por la presencia del médico durante la toma de la tensión arterial y se manifiesta como una reducción de las cifras de tensión arterial de los pacientes hipertensos durante el seguimiento, hasta el punto de dejar de cumplir los criterios de hipertensión<sup>10</sup>. Sin embargo, una parte de los enfermos etiquetados como “hipertensos de bata blanca” no son sino sujetos erróneamente clasificados como hipertensos a consecuencia de la variabilidad de la tensión arterial y de los errores de medida asociados a su determinación. Este hecho justifica la utilización de protocolos rigurosos en el diagnóstico de la hipertensión en general, y de la hipertensión de bata blanca en particular<sup>11-13</sup>.

Así, en un estudio danés se evaluó la proporción de hipertensos de bata blanca en 153 hipertensos de nuevo diagnóstico con una presión diastólica mayor de 90 mmHg<sup>14</sup>. Estos pacientes fueron remitidos a una clínica de hipertensión donde se realizaron determinaciones más precisas, encontrándose que 28 (18,3%) presentaban cifras medias de presión diastólica menores de 90 mmHg. Los autores consideraron que estos 28 pacientes tenían hipertensión de bata blanca. No obstante, las cifras inicialmente elevadas y posteriormente normales de una parte de estos casos son consecuencia, probablemente, de errores de clasificación atribuibles al efecto de la regresión a la media –aunque los autores del artículo no discuten esta posible interpretación.

En el seguimiento ambulatorio posterior de estos 153 pacientes se observó que aún persistían errores de clasificación, ya que de los 125 hipertensos restantes 31 (24,8%) presentaron cifras inferiores a 90 mmHg, mientras que 5 de los 28 pacientes (17,9%) que dejaron de ser hipertensos en la medición de la clínica volvieron a presentar cifras elevadas. Irónicamente, el aparente aumento de tensión arterial de estos 5 casos puede deberse a un fenómeno de “regresión alejándose de la media” debido al sentido de la selección de los 28 pacientes.

La regresión a la media tiene una considerable importancia en la práctica clínica diaria, en la que los pacientes son clasificados con frecuencia en función de una única visita o prueba diagnóstica. La correcta consideración de regresión a la media en el ámbito clínico evitaría en muchas ocasiones la realización de pruebas complementarias adicionales y la instauración de tratamientos innecesarios.

### **Evaluación sesgada de la eficacia de las intervenciones clínicas**

Aunque cada vez son menos frecuentes, todavía aparecen en la bibliografía ensayos clínicos en los que se pretende evaluar la eficacia de una intervención clínica o farmacológica sin la utilización de un grupo control, realizando comparaciones del tipo “antes-después”. Del mismo modo, en ensayos clínicos con grupo control suele analizarse por separado el grupo de pacientes que ha recibido tratamiento, para encontrar que la intervención es efectiva única o principalmente en aquellos que están en peor situación al inicio. En estos estudios, una buena parte de los efectos observados suele deberse al fenómeno de la regresión a la media. De hecho, la existencia de la regresión a la media es una de las prin-

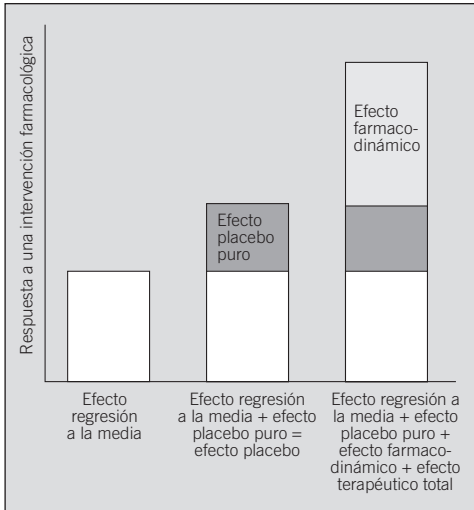


Fig. 2. Efecto de la regresión a la media como componente de la respuesta a placebo y la respuesta global a un fármaco. La respuesta global a una intervención farmacológica tiene un importante componente no farmacodinámico. Los estudios controlados con placebo nos ayudan a eliminar el ruido de fondo producido por el efecto placebo y el efecto de la regresión a la media.

cipales razones por las que muchas intervenciones resultan tan prometedoras en estudios iniciales no controlados y que justifican el establecimiento de un grupo control en la evaluación de las intervenciones clínicas o terapéuticas.

Las consecuencias de la regresión a la media en los ensayos clínicos han sido formalizadas con detalle<sup>2,15</sup>, existiendo abundantes ejemplos de conclusiones erróneas sobre la eficacia de intervenciones terapéuticas debidas a este fenómeno, en particular en enfermedades crónicas que presentan exacerbaciones y remisiones en su evolución.

El proceso de reclutamiento de pacientes para los ensayos clínicos selecciona de forma natural a sujetos con exacerbaciones de su condición clínica que, en muchas ocasiones, tendrán un curso natural hacia la remisión. De la misma manera, los criterios de inclusión/exclusión que se establecen seleccionan a quienes se encuentran por encima (o por debajo) de unos valores que actúan como criterios diagnósticos. En ambos casos, la selección que se precisa para que se produzca la regresión a la media puede ser indirecta, al seleccionarse a los pacientes según algún síntoma o dato de laboratorio correlacionado con la variable de valoración final del estudio. En la evaluación posterior a la intervención, los indi-

viduos que se encontraban en el peor momento de su enfermedad, o que presentaban valores por encima del criterio de selección, tenderán a presentar valores menos extremos, dando la impresión de que la intervención ha sido eficaz<sup>6</sup>.

En un estudio sobre la historia natural de la enfermedad en 556 pacientes con diagnóstico de litiasis biliar sintomática<sup>16</sup>, se observó que, incluso antes de la instauración del tratamiento, los casos más graves tendían a encontrarse mejor y los más leves a encontrarse peor. En el grupo intermedio, aproximadamente la mitad de los pacientes permanecía en la misma situación, mientras que el resto se repartía a partes iguales entre los que empeoraban y los que mejoraban. Esta situación es la esperada si únicamente interviene la regresión a la media. La evaluación de cualquier intervención terapéutica en este contexto puede conducir a conclusiones erróneas.

### Solapamiento con el efecto placebo

En la investigación clínica, la regresión a la media explica, total o parcialmente, la aparente eficacia de muchas intervenciones en ensayos no controlados, la existencia del "efecto placebo"<sup>2,17</sup> (aunque no es infrecuente encontrar monografías sobre el efecto placebo en las que ni siquiera se discute la regresión a la media)<sup>18</sup> y la existencia de subgrupos que parecen responder particularmente bien a una intervención determinada. Así, en un estudio abierto, no controlado, de prevención de las exacerbaciones clínicas en pacientes con bronquitis crónica, no sorprende comprobar que el tratamiento con un mucolítico fue eficaz fundamentalmente en los que tenían exacerbaciones con más frecuencia antes de iniciar el tratamiento<sup>19</sup>. Es interesante notar que pese al elevado tamaño muestral de este estudio (más de 5.000 personas) sus conclusiones no fueron válidas. Tampoco sorprende comprobar cómo, en casos de diverticulosis, el salvado acelera el tránsito intestinal de los pacientes con tiempos de tránsito inicialmente más lento y prolonga los de aquellos con tiempos inicialmente más rápidos<sup>20</sup>. Este mismo fenómeno es un aliado frecuente de los médicos (y más importante todavía, de curanderos y charlatanes), quienes, una y otra vez, comprueban cómo los pacientes que están inicialmente en peor estado son los que más parecen beneficiarse con el tratamiento.

Pero si parece claro que parte del efecto placebo, en muchos ensayos clínicos, puede explicarse por el efecto de regresión a la media, también puede ocurrir que sus efectos se superpongan dando lugar a sorprendentes respuestas al placebo (fig. 2). Esto podría ayudar a explicar la funci-

tuación de la respuesta a placebo que aparece en muchas enfermedades. Por ejemplo, en la úlcera duodenal se han descrito ensayos clínicos con respuesta a placebo hasta en un 48 % de los pacientes incluidos<sup>21</sup>.

### Ineficiencia en la planificación de ensayos clínicos y estudios epidemiológicos

El no considerar la regresión a la media puede tener graves consecuencias en el desarrollo de ensayos clínicos o estudios epidemiológicos. Tal como se ha indicado anteriormente, en la realización de ensayos clínicos es frecuente establecer como criterio de inclusión/exclusión la presencia de valores por encima (o por debajo) de un determinado punto de corte para una variable. Debido al efecto de la regresión a la media, e independientemente de la eficacia de las intervenciones, una parte de los individuos seleccionados en el estudio puede dejar de cumplir los criterios de inclusión, de forma que se estudie la intervención en “pacientes” que realmente no tienen la enfermedad o que la tienen en un grado menor al esperado. Este hecho se va a traducir en una pérdida de la potencia del estudio para detectar efectos, o incluso en la imposibilidad de detectarlos.

Uno de los ejemplos mejor conocidos es el del Multiple Risk Factor Intervention Trial (MRFIT)<sup>22</sup>, un ensayo clínico de prevención primaria en el que se evaluaba el efecto de un programa de intervención multifactorial sobre la mortalidad por enfermedad coronaria en varones de alto riesgo. La selección de los individuos como de alto riesgo, y por tanto susceptibles de ser incluidos en el estudio, se realizó en función de la presencia, en una primera revisión, de niveles elevados de tensión sanguínea, colesterol sérico y consumo de tabaco. Posteriormente, y antes del inicio de la intervención, se repitió la cuantificación de estas variables para utilizarla como medida basal. En el caso de la tensión arterial, la media de presión diastólica de los individuos seleccionados disminuyó desde 99,2 mmHg en la visita de *screening* a 91,0 en la determinación basal (fig. 3), por lo que el número de hipertensos incluidos en el estudio fue muy inferior al previsto, y la potencia del estudio se vio sensiblemente reducida.

El adecuado diseño de los estudios clínicos y epidemiológicos requiere la optimización de los recursos disponibles, tanto materiales como humanos. Olvidar el efecto de la regresión a la media puede encarecer la investigación clínica al tener que aumentar el número de pacientes reclutados, o bien, dar lugar a una incorrecta utili-

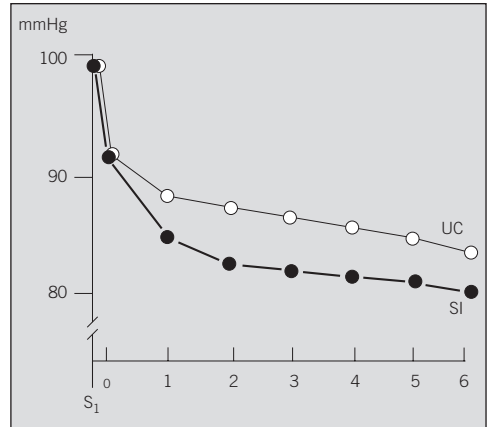


Fig. 3. Evolución de la media de tensión arterial diastólica en los participantes del Multiple Risk Factor Intervention Trial (MRFIT)<sup>22</sup>. Tanto en el grupo de tratamiento habitual (UC) como en el de intervención especial (SI), la tensión arterial diastólica se redujo entre la primera visita de cribado y el valor basal en 8,2 mmHg, un efecto debido, en su mayor parte, a la regresión a la media. En el eje de abscisas se presenta el tiempo de seguimiento de los participantes, en años. (Reproducida con permiso de JAMA 1982; 248: 1.465-1.477. © 1982, American Medical Association.)

zación de los recursos al no poder obtener conclusiones relevantes.

### La regresión a la media como error de medida no diferencial

En los estudios de asociación de un factor de exposición con una variable de efecto, es habitual categorizar los factores de exposición continuos para evaluar su posible papel etiológico y la forma de la curva de dosis-respuesta. Así, por ejemplo, los valores de colesterol en plasma pueden agruparse en “quintiles” (cinco grupos con el mismo número de participantes) con el fin de evaluar el riesgo de enfermedad de unos quintiles frente a otros. Los errores de medida inherentes a la técnica de cuantificación, unidos a la clasificación selectiva en categorías discretas, dan lugar a la presencia de regresión a la media: algunos de los sujetos que se han clasificado en el quintil superior realmente deberían pertenecer a quintiles inferiores y viceversa. Se incurre, por tanto, en una mala clasificación de la exposición, que se manifiesta como una atenuación del efecto estudiado (sesgo hacia el valor nulo) y una pérdida de potencia del estudio. Este efecto es un caso especial de los denominados errores de me-

dida no diferenciales (ya que afectan a la variable de exposición, pero no dependen de la enfermedad desarrollada), y puede abordarse con la metodología epidemiológica y estadística existente para este tipo de sesgos<sup>23</sup>.

En el caso de que la variable de categorización no sea el factor de exposición principal, sino un posible factor de confusión que es necesario controlar, la presencia de regresión a la media impide que el ajuste estadístico sea completo, quedando un cierto grado de confusión residual a pesar de incluir el factor de confusión en los análisis estratificados o modelos estadísticos. En este caso, el sesgo que se introduce en la evaluación del factor de exposición principal puede tener cualquier dirección, llevándonos a sobreestimar o infraestimar el efecto estudiado.

### **¿Cómo controlar el efecto de la regresión a la media?**

La regresión a la media es una fuente potencial de error con importantes consecuencias tanto en la investigación como en la práctica clínica, que ha de ser evitada en la medida de lo posible. Para que se produzca regresión a la media deben coexistir tanto un proceso de selección en función de los valores de una variable, como un cierto grado de variabilidad intraindividual o del método de medición. Son múltiples las situaciones en las que la actuación clínica está restringida a un conjunto de sujetos seleccionados, con valores por encima o por debajo de unas cifras determinadas. Dado que la selección es, en buena medida, inevitable, una estrategia posible de control de la regresión a la media radica en actuar sobre las fuentes de variabilidad intraindividual.

Las fuentes principales de variabilidad intraindividual son la variabilidad fisiológica, los cambios en las condiciones de medida, y los errores de medida aleatorios. La estandarización de la toma de medidas, la utilización de personal entrenado, la certificación y validación de los instrumentos de medida, y el uso de pruebas precisas son buenas prácticas clínicas que contribuyen a disminuir la variabilidad de las determinaciones y, por tanto, el efecto de la regresión a la media. No obstante, la eliminación completa de todas las fuentes de variabilidad es imposible, por lo que se han propuesto diversos métodos para controlar el efecto de la regresión a la media. Su aplicación va a depender del objetivo de la actuación o investigación clínica.

En los ensayos clínicos, la inclusión de un grupo control evita directamente el sesgo en la estimación del efecto que produce la regresión a la

media, ya que tanto el grupo control como el de intervención sufren el mismo grado de regresión y, por tanto, cualquier diferencia entre ambos será atribuible al efecto de la intervención<sup>3</sup>. Hoy día resulta difícilmente justificable la realización de estudios de eficacia o efectividad no controlados.

La existencia de un grupo control, sin embargo, no evita los errores de clasificación y la dilución que éstos producen en las estimaciones de efecto. Una estrategia eficaz consiste en mejorar el proceso de selección tomando como determinación basal para cada paciente la media de una serie de medidas iniciales. Dado que la variabilidad de la media es menor que la variabilidad de cualquier determinación aislada, aquélla será una estimación más precisa del valor “verdadero” de cada sujeto, lo que permite reducir tanto el efecto de la regresión como los errores de clasificación<sup>3,7</sup>. No obstante, la disminución del efecto de la regresión a la media que se consigue con cada nueva medida adicional es progresivamente menor, por lo que el número de medidas utilizado en cada caso depende del grado de regresión admisible y del coste y facilidad de obtención de las mismas<sup>3</sup>.

Una alternativa para controlar tanto el efecto de la regresión como los errores de clasificación consiste en realizar la selección siguiendo una estrategia en dos etapas<sup>4</sup>. En la selección inicial se utilizan criterios más permisivos y un número reducido de determinaciones. Posteriormente se aplican sobre los pacientes seleccionados criterios más estrictos y un mayor número de determinaciones, cuya media será usada como nivel de referencia basal. Esta estrategia produce una disminución importante de los errores de clasificación y del efecto de regresión, reduce la variabilidad basal de la variable de valoración y evita la realización de determinaciones múltiples a todos los pacientes en estudio.

Finalmente, existen métodos estadísticos que permiten, bajo ciertas asunciones, estimar la importancia que la regresión a la media tiene en los resultados obtenidos<sup>5</sup>. Sin embargo, estos métodos tienen un papel secundario, ya que los efectos de la regresión a la media deben controlarse mediante diseños apropiados y la realización de determinaciones clínicas precisas.

En conclusión, la regresión a la media afecta a la mayor parte de las actuaciones clínicas en las que existe un proceso de selección. Sin embargo, sus efectos pueden ser fácilmente minimizados mediante la planificación de actuaciones sencillas en el momento de la selección de los pacientes, evitando la realización de pruebas secundarias no justificadas, la administración de tratamientos in-

necesarios, la obtención de conclusiones erróneas de las investigaciones clínicas o la utilización ineficiente de los recursos materiales y humanos.

### Agradecimiento

Agradecemos a los doctores Javier Cid, Javier Damián, Enrique Gómez-Barrena y Mónica Tafalla sus comentarios de las versiones previas del artículo.

### BIBLIOGRAFÍA

- Galton F. Regression towards mediocrity in hereditary stature. *J Anthropol Inst* 1886; 15: 246-263.
- Andersen B. Methodological errors in medical research. An incomplete catalogue. Oxford: Blackwell Scientific Publications, 1990.
- Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol* 1976; 104: 493-498.
- McMahan CA. Regression toward the mean in a two-stage selection program. *Am J Epidemiol* 1982; 116: 394-401.
- James KE. Regression toward the mean in uncontrolled clinical studies. *Biometrics* 1973; 29: 121-130.
- Fleiss JL. The design and analysis of clinical experiments. Nueva York: John Wiley & Sons, 1982.
- Gardner MJ, Haddy JA. Some effect of within-person variability in epidemiological studies. *J Chronic Dis* 1973; 26: 781-795.
- Ederer F. Serum cholesterol changes: effects of diet and regression toward the mean. *J Chronic Dis* 1972; 25: 277-289.
- Gould BA, Mann S, Davies AB, Altman DG, Rafferty EB. Does placebo lower blood-pressure? *Lancet* 1981; 2: 1.377-1.381.
- Mancia G, Grassi G, Pomidossi G, Gregorini L, Bertinieri G, Parati G et al. Effects of blood-pressure measurement by the doctor on patient's blood pressure and heart rate. *Lancet* 1983; 2: 695-698.
- Rosner B. Screening for hypertension-some statistical observations. *J Chron Dis* 1977; 30: 7-18.
- Shepard DS, Finison LJ. Blood pressure reductions: correcting for regression to the mean. *Prev Med* 1983; 12: 304-317.
- Amigo I, Herrera J. La hipertensión de bata blanca. *Med Clin (Barc)* 1994; 102: 64-66.
- Høegholm A, Kristensen KS, Madsen NH, Svendsen TL. White coat hypertension diagnosed by 24-h ambulatory monitoring. Examination of 159 newly diagnosed hypertensive patients. *Am J Hypertension* 1992; 5: 64-70.
- Spector R, Park GD. Regression to the mean: a potential source of error in clinical pharmacological studies. *Drug Intell Clin Pharm* 1985; 19: 916-919.
- McSherry CK, Ferstenberg H, Calhoun WF, Lahman E, Virshup M. The natural history of diagnosed gallstone disease in symptomatic and asymptomatic patients. *Ann Surg* 1985; 202: 59-63.
- McDonald CJ, Mazzuca SA, McCabe GP. How much of the placebo "effect" is really statistical regression? *Stat Med* 1983; 2: 417-427.
- Spiro HM, ed. Doctors, patients, and placebo. New Haven: Yale University Press, 1986.
- Anónimo. Prevention of chronic bronchitis exacerbations with ambroxol (Mucosolvan Retard). An open, long-term, multicenter study in 5,635 patients. *Respiration* 1989; 55 (Supl 1): 84-96.
- Brodrribb JM, Humphreys DM. Diverticular disease: three studies. Part II - treatment with bran. *Br Med J* 1976; 1: 425-428.
- Elander G, Hermeren G. Placebo effect and randomized clinical trials. *Theor Med* 1995; 16: 171-182.
- Multiple Risk Factor Intervention Trial Research Group. Multiple Risk Factor Intervention Trial. Risk factor changes and mortality results. *JAMA* 1982; 248: 1.465-1.477.
- Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. *Annu Rev Publ Health* 1993; 14: 69-93.

---

# El tamaño del efecto de la diferencia entre dos medias: ¿estadísticamente significativo o clínicamente relevante?

---

Alfonso Casado<sup>a</sup>, Luis Prieto<sup>b,c</sup> y Jordi Alonso<sup>b</sup>

<sup>a</sup>Institut Universitari de Salut Pública de Catalunya. Universitat de Barcelona.

<sup>b</sup>Unitat de Recerca en Serveis Sanitaris. Institut Municipal de Investigació Mèdica (IMIM). Barcelona.

<sup>c</sup>Facultat de Psicologia i Ciències de l'Educació Blanquerna. Universitat Ramon Llull. Barcelona.

El problema de la reflexión lógica y estadística en torno a los datos, junto al de extraer conclusiones a partir de los mismos, forma parte esencial de la investigación en salud. En dos artículos aparecidos en *MEDICINA CLÍNICA*<sup>1,2</sup>, se expone de forma amena y dialogante, poco usual en la bibliografía médica, este tema bajo diversos enfoques de pensamiento. Propone su autor que toda conclusión o proposición del tipo “el tratamiento A es mejor que el B”, resultaría menos informativa que una buena medición de la diferencia entre los tamaños de los efectos (TE) producidos por A y B, y que un hábito inveterado de economía intelectual sería el factor que por inercia mantiene la anterior forma de pensamiento. Sin embargo, el propio autor también reconoce que ya existen probados indicios (siendo el metaanálisis el más elocuente) del desplazamiento del interés de los investigadores hacia la medición de los efectos, con la consecuente tendencia a abandonar el esquema estrecho de la significación estadística.

Efectivamente, basta una lectura rápida de los artículos científicos publicados en las revistas biomédicas, para darse cuenta de que, con frecuencia, la preocupación de sus autores acerca de los efectos encontrados se basa en el enfoque de la inferencia estadística propuesto por Fisher<sup>3</sup>. Este enfoque se centra exclusivamente en el resultado de la prueba estadística calculada, y especialmente en el valor de significación (valor p) asociado a dicha prueba: a partir de la evaluación del valor p, y teniendo en cuenta un criterio arbitrariamente estandarizado (p. ej., valor p mayor o menor o igual que 0,05), se determina si el fenómeno estudiado está presente (la hipótesis nula [ $H_0$ ] es falsa) o ausente (la  $H_0$  es cierta) en la población de la que proviene la muestra estudiada.

Sea cual sea la naturaleza del fenómeno estudiado, aceptar la  $H_0$  siempre significa que el fenómeno no existe. La forma literal en la que la  $H_0$  es enunciada dependerá del tipo de prueba estadística empleada. Por ejemplo, en circunstancias en las que dos poblaciones expuestas a dos tratamientos diferentes son comparadas en cuanto a la media de días de ingreso hospitalario que producen, la  $H_0$  tomará habitualmente la forma de: “la diferencia en la media de días de ingreso entre los dos tratamientos es igual a 0”.

Si enunciamos de forma general el concepto de “TE” como “grado en que el fenómeno estudiado está presente en la población”<sup>4</sup>, la aceptación de la  $H_0$  bajo el enfoque fisheriano siempre significa que el TE es igual a 0, o sea, que no existe el fenómeno en cuestión. Siguiendo con el enfoque de Fisher, el rechazo de la  $H_0$  únicamente implica que el TE puede tomar cualquier valor diferente del 0, consideración que claramente imposibilita el cálculo de un valor exacto para este TE y, por tanto, la cuantificación del efecto producido.

Fueron Neyman y Pearson quienes formularon una alternativa al método de inferencia estadística propuesto por Fisher que sí permite el cálculo exacto del TE<sup>4</sup>. Brevemente, indicaremos que, además de la especificación de una  $H_0$ , el método de Neyman-Pearson propone la existencia de una hipótesis alternativa ( $H_1$ ), contrapuesta a la  $H_0$ . Como estrategia para cuantificar el grado en que el fenómeno estudiado está presente en la población, o dicho de otro modo, para cuantificar el grado en que la  $H_0$  es falsa, se determinará cuál es la discrepancia entre  $H_0$  y  $H_1$ . Esta discrepancia recibe el nombre de TE y es una de las variables que, junto al tamaño de la muestra, el nivel de significación ( $\alpha$ ) y la potencia estadística ( $1-\beta$ ), está implicada en el proceso de inferencia estadística tal y como hoy lo conocemos.

A pesar de la reconocida importancia que el concepto de TE tiene como forma de pensamien-

---

Publicado en *Med Clin (Barc)* 1999; 112 (15): 584-588. Este trabajo ha sido financiado, en parte, por la CIRIT (1997 SGR 00359).

to estadístico actual, y al igual que cualquier otra nueva idea que tenga que abrirse paso enfrentada a la habitual inercia científica: a) su uso sigue apareciendo en pocas publicaciones de salud y con muy baja frecuencia, y b) se enuncia de diversas formas, habitualmente dispersas y no coherentes, y desde varios y no siempre coincidentes enfoques de aplicación y pensamiento teórico<sup>5</sup>.

El presente trabajo pretende sintetizar algunos aspectos básicos relacionados con el concepto de TE, así como los relacionados con su aplicación y utilidad en la investigación clínica.

### El índice del tamaño del efecto (ITE)

Hasta el momento, el TE ha sido considerado de forma abstracta como un parámetro que puede tomar cualquier valor, incluyendo el 0 cuando no existe discrepancia entre  $H_0$  y  $H_1$ . En la práctica, es posible calcular un ITE para cualquier test estadístico. Cohen<sup>4</sup> ha proporcionado formulaciones de cálculo para diferentes pruebas, entre las que destacan: la prueba de la t de Student de comparación de medias, el coeficiente de correlación de Pearson, la prueba de diferencias entre proporciones, la prueba de la  $\chi^2$ , el análisis de regresión múltiples y el análisis de la variancia. Todos los índices propuestos brindan una medida de la magnitud del efecto producido, libre de cualquier unidad de medida específica.

Siguiendo con el ejemplo esbozado con anterioridad y mediante un razonamiento generalizable a otro tipo de análisis, ilustraremos el cálculo del ITE para el caso de una de las pruebas estadísticas más comúnmente empleadas en la bibliografía biomédica: la prueba de la t de Student de comparación de dos medias provenientes de 2 muestras, cada una con n casos, y ambas aleatoria e independientemente extraídas de dos poblaciones normalmente distribuidas. Con dicha prueba, el investigador trata de comprobar la  $H_0$  de que la diferencia entre sus respectivas medias poblacionales es nula. Una forma de cálculo del ITE para la situación planteada se efectúa estandarizando el efecto bruto expresado en las unidades de la variable dependiente, en función de la desviación estándar (común) de las medidas en sus respectivas poblaciones, también expresada en las unidades originales de medida. Para las dos muestras independientes, y en el caso de una prueba bilateral, se realiza según la fórmula:

$$ITE = (|M_A - M_B|) / DE_{A,B}$$

donde  $M_A$  y  $M_B$  designan, respectivamente, las medias de las muestras A y B, y  $DE_{A,B}$ , la desvia-

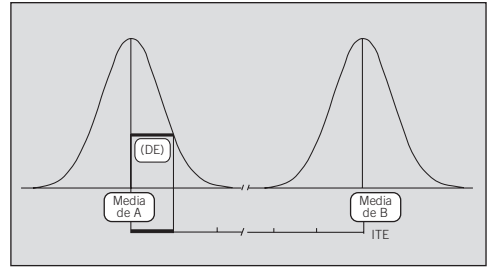


Fig. 1. Representación gráfica del significado del tamaño del efecto (TE) entre dos medias (si, por ejemplo, la media de A es 10, la de B 40 y la desviación estándar [DE] común 2, el índice del tamaño del efecto [ITE] es  $[(10-40)/2] = 15$ ).

ción estándar de A y B, que por el momento, como ya hemos apuntado, supondremos equivalentes. Así definido, el ITE resulta adimensional y expresa la cantidad de unidades de DE que separan a las medias  $M_A$  y  $M_B$  tal y como se demuestra gráficamente en la figura 1.

### Significación estadística contra relevancia clínica

Un ejemplo con datos hipotéticos contribuirá a aclarar los conceptos hasta ahora enunciados: supongamos que se han ensayado dos técnicas quirúrgicas, A y B, para el mismo tipo de dolencia y que se han registrado los días de ingreso hospitalario para cada técnica, encontrándose las cifras enunciadas en la tabla I. El uso de la prueba de la t de Student, calculada, recordemos, para contrastar la hipótesis de la ausencia de una diferencia estadísticamente significativa entre las dos medias, produciría un valor de  $t = -1,36$  con un valor asociado de  $p = 0,093$ . Por lo tanto, y aceptando como "diferencia significativa" aquella que produjera una  $p \leq 0,05$ , resultaría que las técnicas A y B no presentan una diferencia estadísticamente significativa en el número de días de estancia hospitalaria. Sin embargo, una observación que destaca a simple vista es que la media de días de ingreso de los pacientes del grupo B ha sido de 5 días más que los del grupo A, de tal modo que los pacientes del grupo B han hecho uso de 50 días más de ingreso que los pacientes del grupo A. Dado que A ha producido 200 días de ingreso en total, B ha resultado en un 25 % más de días de ingreso que A. El cálculo del ITE en este caso produciría la cifra  $ITE = (25-20)/8,156 = 0,61$ , valor que, según una simple interpretación<sup>4</sup> que luego será



TABLA I  
DÍAS DE INGRESO HOSPITALARIO  
ASOCIADOS A DOS TÉCNICAS  
QUIRÚRGICAS (A Y B) DIFERENTES

Técnica A		Técnica B	
N.º de paciente	Días	N.º de paciente	Días
1	10	11	15
2	10	12	15
3	10	13	15
4	20	14	25
5	20	15	25
6	20	16	25
7	20	17	25
8	30	18	35
9	30	19	35
10	30	20	35
$M_A = 20$ $DE_A = 8,165$		$M_B = 25$ $DE_B = 8,165$	

M: media de los días de ingreso en A y B,  
DE: desviación estándar de los días de ingreso  
en A y B.

explicada más detalladamente, se entendería como una diferencia de magnitud “media” entre los dos grupos. Además de la interpretación cualitativa del valor obtenido para el ITE, el sentido común haría dudar de la ausencia de una diferencia estadística entre ambos grupos dado el reducido tamaño muestral, por lo que seguramente recomendaría una “ampliación” de los efectivos de cada grupo en aras de verificar si esta situación permanece o varía.

Siguiendo con esta línea de razonamiento, supongamos ahora que se realizan 10 intervenciones más con cada una de las técnicas quirúrgicas, y que se obtienen exactamente los mismos resultados que los presentados inicialmente en la tabla I (por tanto, se duplican literalmente los datos de la tabla I). Un nuevo cálculo de los valores de las medias y las DE para los grupos A y B (ahora con 20 pacientes cada uno) conducirá ahora a los resultados  $M_A = 20$ ,  $DE_A = 7,95$ ,  $M_B = 25$  y  $DE_B = 7,95$ . El cálculo de la prueba estadística resultará en un valor de  $t = -1,98$  y  $p = 0,027$ . Ahora, a pesar de que se ha producido el mismo cambio promedio entre las dos técnicas quirúrgicas, al disponer de un tamaño muestral mayor, podemos afirmar que la diferencia entre ambas es estadísticamente significativa. Obsérvese que la significación  $p$  ha disminuido su valor cuatro veces y media (un 450 %) respecto al cálculo anterior; sin embargo el ITE =  $(25-20)/7,95 = 0,63$  ha aumentado su va-

lor en sólo un 2 % respecto al ITE inicial calculado con un tamaño muestral menor. Desde la óptica puramente estadística, la interpretación ha cambiado radicalmente, y ahora se puede decir que ambas técnicas producen estancias hospitalarias significativamente distintas ( $p = 0,027$ ); sin embargo, desde un punto de vista clínico, y tal como el ITE ha puesto en evidencia, el fenómeno estudiado ha permanecido esencialmente invariable (si bien es cierto que la variancia del fenómeno se ha visto modificada).

### La interpretación del ITE

Si bien la interpretación de los resultados de un test de significación es materia reiteradamente tratada en libros y artículos especializados en la bioestadística, la epidemiología o la investigación médica, no sucede lo mismo con la interpretación de las medidas del TE. La interpretación más sencilla de cualquier ITE asociado a una prueba estadística se basa en una convención propuesta por Cohen<sup>4</sup> centrada en la proposición de unos valores (puntos de corte) que sirvan como definiciones operacionales de los adjetivos “pequeño”, “mediano” y “grande”. Para el caso que nos ocupa, el ITE de la diferencia entre dos medias, Cohen ha propuesto valores de 0,2, 0,5 y 0,8, respectivamente. Si bien es evidente que los valores asociados a estos adjetivos son totalmente relativos y que su significado depende específicamente del contexto de la investigación desarrollada, su uso puede ser justificado en la medida en que cualquier otra convención (p. ej., el nivel de significación de  $p$  en 0,05) pueda serlo. Siempre que no se abuse de su interpretación, lo único que debemos pedirle a una convención arbitraria es que sea razonable. Cohen<sup>4</sup> ha justificado estos puntos de corte a partir de su experiencia con los ITE y también, como veremos a continuación, a partir de otras interpretaciones más elaboradas.

Para el ejemplo que discutimos, Cohen<sup>4</sup> ha propuesto tres opciones adicionales de interpretación, fundamentadas todas ellas en la asunción de que las poblaciones comparadas mediante la prueba de la  $t$  de Student son normales, igualmente numerosas y de igual variabilidad (DE). Las tres interpretaciones se centran en el ITE como medida del porcentaje de no solapamiento de las distribuciones comparadas.

La primera interpretación (I1) se basa en el cálculo del porcentaje no solapado del área cubierta por las dos poblaciones; en la figura 2, sería equivalente al cociente que formaría la medida del área sombreada dividida entre el área total ence-

rrada por las curvas. Para el ejemplo que nos ocupa y para entender el significado de I1, quizá sea mejor señalar que el área solapada (1-porcentaje no solapado) nos proporciona información sobre la proporción de sujetos cuya cantidad de días de ingreso pudo haberse producido tanto bajo el uso de la técnica A como de la B, de tal manera que a mayor ITE, menor porcentaje solapado.

La segunda interpretación (I2) se basa en el porcentaje del área comprendida (con relación al área total bajo la curva) entre el punto de intersección de las dos curvas de distribución y el extremo de la cola opuesta para una cualquiera de las distribuciones (fig. 3). En ésta, la zona sombreada indicaría el porcentaje de sujetos de la intervención B que superarían a idéntico porcentaje de sujetos en la intervención A.

Bajo la tercera interpretación (I3) del ITE (fig. 4), el porcentaje del área comprendida en A, desde su cola izquierda, hasta la media de B, nos indicaría el porcentaje de sujetos de A que son superados por el 50% de los sujetos de B.

Cualquiera de estas interpretaciones puede ser utilizada independientemente. Si recordamos que el valor del ITE para la prueba de la t de Student está expresado en unidades de desviación bajo la curva normal (z) y definimos  $P_{ITE}$  como el área (porcentaje de casos) bajo la curva a la izquierda de un valor de ITE (z) dado, el cálculo de I1, I2, e I3 para un ITE cualquiera se puede realizar de forma sencilla a partir de una tabla de z y considerando las siguientes fórmulas<sup>4</sup>:

$$I3 = P_{ITE}$$

$$I2 = P_{ITE/2}$$

$$I1 = (2P_{ITE/2} - 1) / P_{ITE/2}$$

Para calcular I1, I2 e I3 para el ITE = 0,63 obtenido en el segundo caso del ejemplo presentado, sólo tenemos que buscar en una tabla de z el valor z = 0,63 (el ITE) y comprobar que equivale a una proporción ( $P_{ITE}$ ) del 73,57%. Ya sabemos que I3 es igual al 73,57%, o dicho de otro modo, que el 50% de los sujetos de la técnica quirúrgica B superan en tiempo de ingreso hospitalario a un 73,57% de los sujetos intervenidos con la técnica A. A partir de este resultado y sustituyendo en las fórmulas anteriores, podremos comprobar que I2 = 62,36%, o sea que el 62,36% de los sujetos intervenidos con la técnica B permanecen más tiempo ingresados que el 62,36% de los sujetos intervenidos con la técnica A y que I1 = 39,65%, o dicho de forma equivalente, que el 60,35% (1-39,65%) del total de los días de ingreso pudo haberse producido tanto por el uso de la técnica A como de la B.

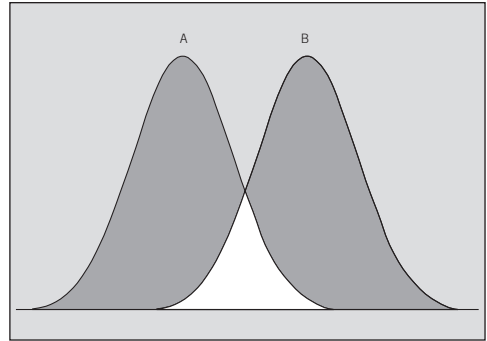


Fig. 2. Representación gráfica de la primera interpretación (I1) del índice del tamaño del efecto (ITE) de la diferencia entre dos medias.

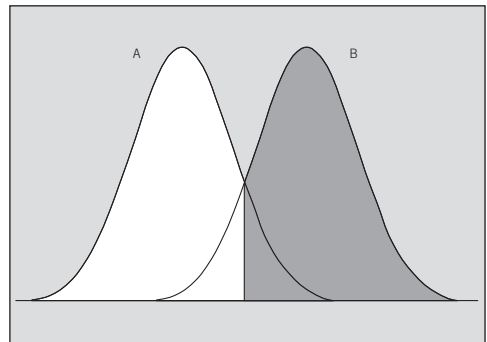


Fig. 3. Representación gráfica de la segunda interpretación (I2) del índice del tamaño del efecto (ITE) de la diferencia entre dos medias.

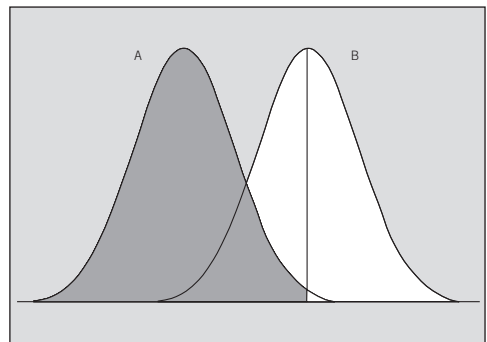


Fig. 4. Representación gráfica de la tercera interpretación (I3) del índice del tamaño del efecto (ITE) de la diferencia entre dos medias.

Empleando la I1, que nosotros calificaríamos como la más fácil de comprender para una persona ajena a la metodología estadística, y enla-

zando con la interpretación arbitraria antes discutida, podemos observar que los ITE de tamaño “pequeño” (0,2), “mediano” (0,5) y “grande” (0,8) equivalen, respectivamente, a un 14,7, un 33 y un 47,4 % de no solapamiento de las áreas combinadas. Cuanto mayor es el ITE, menor es el área de solapamiento entre las dos distribuciones comparadas y esto ilustra que tales puntos de corte arbitrarios responden satisfactoriamente a las exigencias de plausibilidad antes planteadas.

### El incumplimiento de las condiciones de aplicación

La fórmula para obtener el ITE seleccionado como ejemplo, así como las estrategias de interpretación de su resultado, ha sido diseñada para el caso de una prueba t de Student entre dos medias de dos muestras independientes (A y B) (condición I), de igual tamaño (condición II), que provienen de poblaciones de igual variancia (condición III) y normalmente distribuidas (condición IV)<sup>4</sup>. En la práctica puede ocurrir que estas condiciones no se cumplan, por lo que habrá que pensar en formulaciones alternativas adecuadas para cada circunstancia.

El incumplimiento de la condición II, la que supone la igualdad de los tamaños muestrales ( $n_A$  y  $n_B$ ), no afectará a la validez de la interpretación del ITE en términos de I1, I2 e I3 siempre que la DE sea igual en las dos muestras, ya que en tal caso seguiremos asumiendo que el tamaño de las dos poblaciones es el mismo.

En el caso de que las dos DE sean diferentes en las dos muestras, nos encontraremos ante el incumplimiento de una de las asunciones primordiales (condición III), bajo las cuales la prueba t de Student debe ser aplicada. Si a pesar de ello, el tamaño muestral es igual en ambas muestras, existen evidencias de que la prueba de la t de Student mantendrá una moderada firmeza en sus resultados<sup>4</sup>. De todas formas, cuando las DE sean diferentes, el ITE para esta prueba deberá ser modificado. En lugar de utilizar la DE común (ya no lo es) en el denominador de la fórmula propuesta, se deberá obtener una estimación conjunta de las desviaciones ( $DE'$ ) a partir del cálculo de la raíz cuadrada de la media de las dos variancias.

$$DE' = (IDE_A^2 + IDE_B^2/2)^{1/2}$$

Las interpretaciones del ITE obtenido a partir de  $DE'$  mediante la estrategia de calcular I1, I2 e I3 ya no tendrá validez. Sin embargo, se puede mantener la aproximación basada en los puntos de corte 0,2, 0,5 y 0,8 propuesta por Cohen<sup>4</sup>.

Si en lugar de disponer de dos muestras independientes (condición I), lo que se pretende es establecer el TE de las diferencias entre dos medias que provienen de un grupo de n observaciones apareadas, por ejemplo entre evaluaciones clínicas antes de una intervención X y después de ella Y, se utilizará la prueba de la t de Student para datos apareados. En este caso, la prueba de la t de Student de la diferencia entre la media de X y la media de Y se calculará a partir del análisis de la variable diferencia ( $|X-Y| = Z$ ), por lo que la estimación del ITE para este caso particular se llevará a cabo a partir de la siguiente fórmula<sup>4</sup>:

$$ITE' = |m_z|/DE_z$$

en la que  $m_z$  es (o representa) a la media de la diferencia entre X e Y, y  $DE_z$  a la desviación estándar de dicha diferencia. El  $ITE'$  está expresado, en este caso, en unidades de DE de las diferencias; para interpretar sus valores de la misma forma que hemos enunciado para una comparación de muestras independientes, bastará con multiplicarlos por la raíz cuadrada de 2, para compensar la disminución del error de muestreo obtenida al aplicar un diseño apareado.

El incumplimiento de la cuarta condición, que asume la normalidad de las distribuciones implicadas en el análisis de las diferencias, supondrá, en cualquier caso, la imposibilidad de calcular el ITE mediante la formulación aquí propuesta. Se podría sugerir una alternativa, centrada en el uso de las medianas de las distribuciones y del rango intercuartil, en sustitución de las estimaciones de las medias y la DE de la población, respectivamente. Sin embargo, esta nueva aproximación tampoco permitiría una interpretación precisa de los valores de los ITE obtenidos, ya que el cálculo del solapamiento de las distribuciones sólo se puede llevar a cabo asumiendo la normalidad de las mismas.

Una aproximación diferente para solventar el problema de la no normalidad podría centrarse en el desarrollo de un algoritmo que permitiese estimar el porcentaje de solapamiento de las áreas de las distribuciones de los datos, sin necesidad de asumir una forma precisa para ellas, o bien suponiendo alguna otra forma de distribución no normal, convenientemente justificada. Cabe notar que bajo este enfoque lo esencialmente importante no sería el cálculo del ITE, sino su interpretación, valorando directamente el impacto real del “desconocido” TE mediante estrategias como I1, I2 e I3. Esta proposición pudiera ser, desde el punto de vista estrictamente teórico, discutible; sin embargo, desde el punto de vista práctico, su

implementación llenaría un hueco que por vacío nos lleva a dejar análisis por hacer, o bien a aplicar inadecuadamente los métodos descritos, con las imprevisibles consecuencias de esta acción.

## Conclusión

A la hora de analizar los datos obtenidos en el ámbito de la investigación en salud, todavía se enfatiza, a pesar de las críticas suscitadas<sup>1,2,6</sup>, la comprobación de la significación estadística de los resultados. La confianza depositada en esta aproximación tradicional, basada en la estadística inferencial propuesta por Fisher, puede conducir a la aceptación de la  $H_0$  (se acepta que el fenómeno no existe), cuando ésta es en realidad falsa.

El uso de los indicadores puntuales del TE propuestos por Cohen, y asentados en la concepción de la inferencia estadística de Neyman y Pearson, puede subsanar los errores de interpretación suscitados por la falta de potencia en las pruebas estadísticas calculadas, y se convierte en una alternativa complementaria a considerar a la hora de extraer conclusiones a partir de los datos obtenidos.

Para el caso particular de la diferencia entre dos medias, generalizable a cualquier otra prueba estadística, las estrategias de interpretación de los ITE constituyen también una nueva opción para evaluar e identificar cambios clínicamente relevantes, independientes de la unidad de medida considerada y del tamaño de las muestras estudiadas. En este sentido, el cálculo de los ITE ha sido propuesto, por ejemplo, como sistema para valorar si determinados cambios detectados en el estado de salud de los individuos, potencialmente inducidos por una intervención (p. ej., un ensayo clínico), pueden ser considerados clínicamente relevantes<sup>5,7-10</sup>.

Si bien es importante recordar que existe una serie de asunciones básicas a cumplir antes de calcular cualquier ITE<sup>4</sup>, desde aquí quisiéramos

proponer su uso generalizado en aras de una mejor y más adecuada comprensión de los resultados obtenidos a partir del análisis estadístico de los datos.

## BIBLIOGRAFÍA

1. Bacallao J. La perspectiva exploratorio-confirmatoria en las aplicaciones biomédicas de la estadística: dos diálogos (I). Bayesianismo frente a frecuentismo: sus respectivas implicaciones prácticas en relación con el análisis de datos. *Med Clin (Barc)* 1996; 107: 467-471.
2. Bacallao J. La perspectiva exploratorio-confirmatoria en las aplicaciones biomédicas de la estadística: dos diálogos (y II). Consideraciones críticas acerca de las pruebas de significación. *Med Clin (Barc)* 1996; 107: 539-543.
3. Fisher RA. *The design of experiments* (8ª ed.). Edimburgo, Scotland: Oliver & Boyd, 1966. (Trabajo original publicado en 1935.)
4. Cohen J. *Statistical power analysis for behavioral sciences*. Nueva York, NY: Academic Press, 1988.
5. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997; 50: 869-879.
6. Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika* 1928; 20: 175-240.
7. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27 (Supl): 178-189.
8. Deyo RA, Diehr P, Patrick D. Reproducibility and responsiveness of health status measures. *Control Clin Trials* 1991; 12 (Supl): 142-158.
9. Testa MA, Anderson RB, Nackley JF, Hollenberg NK, and The Quality of Life Hypertension Study Group. Quality of Life and antihypertensive therapy in men. A comparison of captopril with enalapril. *N Engl J Med* 1993; 328: 907-913.
10. Cassard SD, Patrick DL, Damiano AM, Legro MW, Tielsch JM, Diener-West M et al. Reproducibility and responsiveness of the VF-14. *Arch Ophthalmol* 1995; 113: 1.508-1.513.

## Parte 3

---

# **CALIDAD Y BUENAS PRÁCTICAS CLÍNICAS**

---

---

# Evaluación crítica de ensayos clínicos

---

J.A. Sacristán, J. Soto<sup>a</sup> e Inés Galende Domínguez<sup>b</sup>

Departamento de Investigación Clínica, Lilly S.A. <sup>a</sup>Unidad de Farmacología Clínica. Hospital Santa Cruz. Liencres. Cantabria.  
Centro Regional de Farmacovigilancia de Cantabria. <sup>b</sup>Subdirección General de Evaluación de Medicamentos.  
Dirección General de Farmacia y Productos Sanitarios. Ministerio de Sanidad y Consumo.

A pesar de sus limitaciones, el ensayo clínico controlado (ECC) continúa siendo la mejor herramienta para evaluar la eficacia y seguridad de un nuevo tratamiento. Para que haya llegado a convertirse en el estándar de oro de la investigación clínica ha sido necesario el análisis minucioso de los defectos y los sesgos existentes en los primeros diseños y la puesta en marcha de medidas encaminadas a evitarlos. De esa forma se han ido desarrollando métodos de evaluación más objetivos, se ha generalizado el empleo de la aleatorización, del enmascaramiento, y del correcto análisis estadístico. Esos son los pilares sobre los que descansa la calidad de un ensayo clínico (EC) y los que garantizan la fiabilidad de los resultados obtenidos<sup>1-5</sup>.

Debido al creciente número de EC que se publican actualmente, puede ser de gran utilidad contar con métodos que ayuden a evaluar de forma rápida y sistemática su calidad, mediante el análisis de los requisitos metodológicos y éticos necesarios para que puedan aceptarse sus conclusiones. Diversas revistas médicas internacionales han analizado los defectos metodológicos más comunes, que aparecen en los ensayos clínicos publicados<sup>6,7</sup>, y en los últimos años han surgido distintos cuestionarios o *check-lists*<sup>8-14</sup>, e incluso algunas revistas, como *British Medical Journal* han propuesto sus propios *check-lists* sobre los apartados que debe tener un ensayo clínico para que pueda ser aceptada su publicación<sup>16</sup>.

Además de su objetivo fundamental, valorar si los métodos utilizados y los resultados obtenidos en la investigación son adecuados para producir información útil<sup>9</sup>, los *check-lists* cumplen una importante labor formativa, aportando información, métodos e hipótesis nuevas y fomentando el sentido crítico con los estudios que ya se han realizado o que se puedan realizar en un futuro<sup>1</sup>.

El objetivo de este trabajo ha sido elaborar unos *check-lists* prácticos, que puedan ser utilizados de forma fácil y rápida para evaluar ensayos clínicos publicados y protocolos de ensayos clínicos. La distinta perspectiva de los autores, que trabajan en áreas complementarias relacionadas con la investigación clínica (administración sanitaria, hospital e industria farmacéutica), y la utilización de los términos en español, son las dos principales ventajas que pueden aportar los *check-lists* que se presentan; sin olvidar que la unidad de criterios a la hora de leer críticamente los estudios publicados, de aceptar trabajos por los editores de revistas y de diseñar y evaluar los protocolos de los ensayos, puede contribuir al aumento de la calidad de la investigación clínica, gracias a la generación de información cada día más válida.

La utilidad fundamental del primero de los *check-lists* propuestos es la evaluación de ensayos clínicos publicados. Su uso ayuda a conocer cuáles son los puntos que se deben analizar y las repercusiones que pueden tener las deficiencias encontradas en los mismos. El segundo *check-list*, muy similar al anterior, sirve para evaluar la calidad en el diseño de protocolos de ensayos clínicos. Puede servir de guía para aquellos investigadores que quieran diseñar un estudio teniendo en cuenta los aspectos metodológicos y éticos que se deben considerar cuando se va a realizar una investigación en seres humanos. También puede ser útil para los Comités Éticos de Investigación Clínica, responsables en gran medida, especialmente cuando se desarrolle la Ley del Medicamento<sup>17</sup>, de garantizar la protección de los sujetos que participan en el ensayo, mediante un riguroso estudio de los aspectos anteriormente señalados.

Tras realizar una revisión de algunos de los *check-lists* publicados<sup>8-15</sup>, se definió un modelo "ideal", que tuviera unas determinadas características generales: extensión, método de evaluación y formato, entre otros. Posteriormente, cada

uno de los autores elaboró de forma independiente sendos cuestionarios, que después fueron confrontados, confeccionando de forma consensuada los definitivos.

Para intentar evitar confusiones, siempre que fue posible se definieron los términos empleados, y se señalaron sus equivalentes en inglés, por el gran arraigo que éstos tienen.

### **Contenido de los cuestionarios (*check-lists*)**

En el anexo I se muestra el *check-list* para la evaluación de ensayos clínicos publicados; y en el anexo II el correspondiente a la evaluación de protocolos de EC.

En ambos casos el formato es similar: se indican unos apartados generales que se deben evaluar (enmascaramiento o métodos estadísticos, etc.), y en cada uno de ellos se formulan las preguntas que deben considerarse para poder valorar de forma correcta cada uno de estos apartados. Por ejemplo, en el apartado "Análisis estadístico" es preciso comprobar si se ha calculado de forma correcta el tamaño de la muestra, o si se indican las pruebas estadísticas utilizadas. A estos "subapartados" se les califica como adecuado (+), dudoso ( $\pm$ ), inadecuado (-) o no aplicable (NA). Tras evaluar esos puntos, y de acuerdo a los mismos, debe realizarse una valoración global del apartado general, calificándolo como correcto (C), aceptable (A), dudoso (R), incorrecto (I) o no aplicable (NA), siempre teniendo en cuenta la particularidad de cada ensayo (fase de desarrollo del fármaco, enfermedad en estudio, o alternativas terapéuticas). Aunque pueden utilizarse para cualquier tipo de ensayo clínico, estos cuestionarios están enfocados de forma especial a la evaluación de ensayos clínicos en fases III y IV. En los ensayos clínicos en fases I y II a veces no será necesario incluir alguno de los apartados (p. ej., cálculo del tamaño de la muestra o enmascaramiento) y quizá sea necesario tener en cuenta otro tipo de consideraciones. Lo mismo ocurre con determinadas enfermedades, que pueden requerir diseños específicos y cuya evaluación mediante estos *check-lists* puede resultar complicada (p. ej., oncología, sida, etc.).

A continuación se definen los términos empleados y se analizan los subapartados que consideran:

*Check-list para evaluación de ensayos clínicos publicados* (anexo I).

*Objetivo del estudio.* Todo ensayo clínico debe estar justificado y orientado hacia un objetivo

principal, correctamente definido y basado en una hipótesis previa plausible. Sólo sabiendo qué es exactamente lo que se busca se puede interpretar de forma adecuada un resultado. En general, el diseño debe estar encaminado a contestar un sola pregunta. Los estudios que pretendan dar respuestas a varias preguntas suelen ser excesivamente complicados y responden habitualmente de forma parcial a todas ellas.

*Criterios de selección de los sujetos.* Deben especificarse los criterios que deben cumplir los posibles sujetos participantes para su inclusión en el estudio: edad, sexo, origen del paciente (hospitalizado, ambulatorio) y la definición del/los criterio/s diagnóstico/s (subtipo, escala de gravedad, etc.), a ser posible de acuerdo a criterios internacionalmente aceptados<sup>10</sup>. Así mismo, se indicarán aquellos criterios que, de presentarse, excluirían a los sujetos, tanto para comenzar el estudio, como caso de aparecer a lo largo del mismo.

Los resultados sólo podrán extrapolarse (representarán) a aquella población de pacientes que cumplan los criterios de selección exigidos a la muestra que participa en el ensayo. Por ello los estudios de tipo explicativo, con unos criterios muy restrictivos, intentan dar respuesta a un aspecto muy concreto de un pequeño grupo de pacientes, y los estudios pragmáticos, con criterios de selección más amplios, darán lugar a resultados que serán extrapolables a la gran mayoría de los pacientes.

*Definición del tratamiento experimental y control.* Para ser reproducible es preciso especificar la dosis, intervalo, duración, etc.<sup>10</sup>. Es necesario valorar si el tratamiento elegido como control es adecuado para el propósito del estudio y si es aceptable desde el punto de vista ético el uso de placebo (en caso de emplearse). También debería señalarse si se permite o no el uso de medicación concomitante, si se valora su influencia o los casos en los que está previsto el uso de medicación de rescate; si se ha realizado un período de lavado (*washout period*), entendiéndose como tal el tiempo necesario para eliminar la influencia de un tratamiento previo (al inicio del ensayo o durante el mismo en diseños cruzados)<sup>18,19</sup>.

Cada vez se está concediendo una mayor importancia a la valoración del cumplimiento terapéutico<sup>10,20</sup>. Es imposible establecer conclusiones si los pacientes no toman correctamente la medicación. Es preciso conocer el grado de cumplimiento y señalar el método empleado para ello: recuento de la medicación sobrante, medida de

ANEXO I  
CHECK-LIST PARA EVALUAR ENSAYOS CLÍNICOS PUBLICADOS

			C A D I N A
<b>1. Definición del objetivo del estudio:</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
¿Existe un objetivo fundamental? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se define una hipótesis <i>a priori</i> ? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		C A D I N A
<b>2. Criterios de selección de los pacientes:</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
¿Se especifican adecuadamente los criterios de inclusión? (edad, sexo, criterios diagnósticos, etc.) .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se especifican adecuadamente los criterios de exclusión? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		C A D I N A
<b>3. Definición del tratamiento experimental y control:</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
¿Es reproducible? (dosis, intervalo, duración, vía de administración, etc.) .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Es el control adecuado? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se especifica el tratamiento concomitante? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Sería necesario un período de lavado? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se mide el cumplimiento y se especifica el método? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
<b>4. Comparabilidad de los grupos de tratamiento (asignación aleatoria):</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	C A D I N A
¿Se realiza asignación aleatoria de los sujetos (aleatorización) .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se especifica el método de aleatorización? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Es la aleatorización ciega? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Son los grupos comparables al inicio? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se necesitaría un período de preinclusión? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
<b>5. Medición objetiva de los resultados (enmascaramiento):</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	C A D I N A
¿Es adecuado el método de enmascaramiento? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Estudio no enmascarado? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Simple ciego? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Doble ciego o <i>double dummy</i> ? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se ha valorado la eficacia del enmascaramiento? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		C A D I N A
<b>6. Variable principal de valoración (end point):</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
¿Es objetiva? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Es relevante desde el punto de vista clínico? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		C A D I N A
<b>7. Criterios de evaluación de la respuesta:</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
¿Es/son objetivo/s? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Es/son sensible/s y específico/s? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Es/son un/os método/s validado/s? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se especifican los criterios de fracaso? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Es suficiente el período de seguimiento postratamiento? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		C A D I N A
<b>8. Pérdidas:</b> .....	+ ± - NA .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
¿Se especifican las pérdidas prealeatorización? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Se especifican las pérdidas postaleatorización? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Abandonos ( <i>drop out</i> )? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Retiradas ( <i>withdrawal</i> )? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
¿Pueden reemplazarse? .....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		

(Continúa)



ANEXO I  
CHECK-LIST PARA EVALUAR ENSAYOS CLÍNICOS PUBLICADOS (Continuación)

	+	±	-	NA		C	A	D	I	NA
<b>9. Diseño estadístico:</b> .....					.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se realiza un cálculo <i>a priori</i> del número de pacientes? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se tiene en cuenta el error alfa? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se tiene en cuenta el poder estadístico? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se tienen en cuenta las diferencias clínicamente relevantes? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se tienen en cuenta los abandonos? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se especifican y son adecuados los test estadísticos empleados? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
<b>10. Análisis de los resultados:</b> .....					.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se realiza análisis por intención de tratar? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se realiza análisis por tratamiento? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se realiza análisis de subgrupos sobre hipótesis <i>a priori</i> ? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Se valora la repercusión de las pérdidas? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
<b>11. Conclusiones:</b> .....					.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Son justificadas (validez interna)? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Puede generalizarse (validez externa)? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
¿Son justificadas las conclusiones por subgrupos? .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						
<b>12. ¿Es adecuado el método de recogida e imputabilidad de acontecimientos adversos?</b> .....						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>13. ¿Se especifica si el ensayo se realizó según BPC?</b> .....						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>14. ¿Se aprobó el protocolo por un Comité ético de investigación clínica?</b> ...						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>15. ¿Se solicitó consentimiento informado a los sujetos participantes en el EC?</b>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

+ = adecuado; ± = dudoso; - = inadecuado; NA = no aplicable; C = correcto; A = aceptable; D = dudoso; I = incorrecto; NA = no aplicable.

valores plasmáticos o urinarios, marcadores químicos o métodos microelectrónicos. Además, es necesario evaluar la repercusión del incumplimiento en función del tipo de tratamiento, por ejemplo es distinto no tomar una dosis de un tratamiento antibiótico, que no tomar una dosis de un tratamiento hipolipemiente.

*Comparabilidad de los grupos de tratamiento.* Para evaluar la respuesta a un determinado factor (tratamiento), es preciso que los grupos en estudio sean comparables en todas sus características excepto en dicho factor. Para intentar conseguir este objetivo es necesario evitar los sesgos de selección, tanto conscientes como inconscientes, y ello se consigue mediante la asignación aleatoria de los sujetos a los grupos de tratamiento<sup>10</sup>. La asignación aleatoria o aleatorización es el procedimiento sistemático y reproducible por el que los sujetos participantes en un estudio clínico son distribuidos al azar en los distintos grupos

de tratamiento, de tal forma que todos ellos tienen la misma probabilidad de recibir cualquiera de los posibles tratamientos<sup>19</sup>. Es un método imprescindible para reducir los sesgos de selección de los pacientes<sup>9,21-24</sup>.

En el caso de factores pronósticos conocidos puede recurrirse a la estratificación, que consiste en la asignación de los pacientes a unos grupos determinados según estos factores pronósticos o características especiales que pueden influir en los resultados del estudio<sup>19,24</sup>. Una vez estratificados (normalmente según dos o tres variables), los pacientes se aleatorizan a uno u otro grupo de tratamiento. Esta técnica es más importante cuando el tamaño de la muestra es inferior a 50 pacientes<sup>19,25</sup>.

En los EC publicados debe comprobarse si los grupos eran comparables al inicio<sup>10</sup>, también mediante estratificación o emparejamiento de los factores de confusión y/o pronósticos más importantes: estado de la enfermedad, enferme-

dades asociadas, entre otros. Es conveniente presentar esos datos de forma tabulada en el apartado de resultados. La aleatorización debe ser ciega<sup>10</sup>, esto es, que el investigador desconozca a qué grupo de tratamiento corresponderá el siguiente paciente. Debería especificarse el método empleado para mantener el ciego (p. ej., sobres cerrados)<sup>19</sup>. En determinados ensayos será necesario realizar un período de preinclusión (*run-in period*), que es el tiempo previo a la aleatorización necesario para estabilizar la situación clínica de los pacientes o valorar el cumplimiento.

*Medición objetiva de los resultados.* Aunque la asignación aleatoria disminuye el sesgo de selección de los pacientes, los resultados del estudio pueden verse distorsionados cuando el paciente o el investigador conocen el tratamiento asignado. Estos sesgos de valoración pueden reducirse mediante el enmascaramiento de las sustancias empleadas en el estudio<sup>19</sup>.

Aunque no debería emplearse, se utiliza el término ensayo abierto o no ciego para los ensayos clínicos en los que tanto el sujeto como el investigador conocen el grupo de tratamiento al que aquél ha sido asignado; simple ciego cuando el sujeto (o a veces el investigador) desconoce el grupo de tratamiento al que pertenece, y doble ciego cuando tanto el paciente como el investigador desconocen la asignación a los grupos de tratamiento<sup>18,19</sup>. Además, cuando se recurre a una tercera persona que desconoce el tratamiento que está recibiendo el sujeto, para evaluar la respuesta, se denomina evaluación ciega por terceros. El doble enmascaramiento (*double dummy*) es un tipo de doble ciego en el que, por resultar imposible lograr que la medicación de los grupos sea idéntica, junto al tratamiento experimental, se administra placebo del tratamiento control y viceversa. Es muy importante que se describan correctamente todas estas medidas en la sección material y métodos del estudio.

La otra pregunta que debe realizarse es si el enmascaramiento ha sido eficaz. A pesar de que el diseño del estudio sea doble ciego, es posible que a lo largo del estudio pueda descubrirse a qué grupo de tratamiento pertenece, bien por presentar efectos adversos predecibles, por aparición de alteraciones de los parámetros de laboratorio, interacciones farmacocinéticas, coloración de la orina.

A veces es imposible realizar un enmascaramiento perfecto de los tratamientos. En tales casos habrá que considerar la repercusión que

puede tener sobre los resultados obtenidos. Tratar de conseguir un enmascaramiento perfecto nunca debe suponer un aumento del riesgo para los sujetos.

*Variable principal de valoración (end point).* En todos los ensayos clínicos debe existir una variable que sea considerada como la más importante para la evaluación final de los resultados. Es deseable que sea lo más objetiva posible y que tenga relevancia desde el punto de vista clínico<sup>19</sup>. No es raro ver ensayos clínicos donde se demuestra una diferencia significativa de una serie de parámetros secundarios, sin ningún interés en la práctica clínica. Esto debe hacer sospechar un análisis *a posteriori* de dichos parámetros, tras no encontrar diferencia significativa en la variable principal de valoración entre el tratamiento experimental y el control.

*Criterios de evaluación de la respuesta.* El criterio de medida debe ser objetivo, lo que es especialmente importante en el caso de estudios multicéntricos, donde los distintos puntos de vista de los investigadores podrían introducir importantes sesgos (principalmente de valoración), llegando a invalidar los resultados<sup>26</sup>.

Es necesario elegir métodos validados, sensibles y específicos<sup>9</sup>, requisitos indispensables para que podamos detectar las posibles diferencias existentes, y asegurar que esos resultados son fiables. En el caso de determinaciones de laboratorio, es necesario indicar que se realizan controles (internos y externos) de calidad en los aparatos, así como los coeficientes de variación detectados.

Los criterios de fracaso deben estar siempre especificados, y el seguimiento postratamiento tiene que ser suficientemente largo, en función de la patología y las características específicas del tratamiento en estudio.

*Pérdidas.* Las pérdidas prealeatorización se refieren a los pacientes elegibles que, por voluntad propia o del investigador, no participan en el ensayo. Deben conocerse las características de estos pacientes, puesto que si difieren de las de los sujetos incluidos, no será posible generalizar los resultados (la muestra sólo será representativa si se aleatorizan todos o la mayoría de los pacientes considerados susceptibles de elección)<sup>19,27</sup>.

También hay que indicar las pérdidas después de la aleatorización<sup>10,19,27</sup>: abandonos (*drop out*) o pacientes que no quieren o no pueden continuar en el estudio y de los que no se puede re-

lizar seguimiento, y las retiradas (*withdrawal*) o pacientes que deben abandonar por aparición de reacciones adversas o de criterios de exclusión secundarios, pero de los que se puede realizar un seguimiento. Ninguno de estos pacientes debería ser sustituido, ya que el nuevo paciente no tendría la misma oportunidad de ser incluido en los grupos del estudio. Para evitar que la pérdida de sujetos disminuya el número global de los mismos y, por tanto, el poder estadístico del estudio, habría que preverlos de antemano e incluirlos al realizar el cálculo del tamaño de la muestra.

*Diseño estadístico.* El diseño estadístico incorrecto es uno de los problemas más comunes en los ensayos clínicos publicados<sup>24,28</sup>, y hay algunos *check-lists* que sólo analizan este aspecto<sup>12,29</sup>. Es esencial realizar un cálculo *a priori* del número de pacientes necesarios para detectar una determinada diferencia, basándose en el error alfa (normalmente de un 5%), en el poder estadístico (o en el error beta) y en lo que se considerará diferencia clínicamente relevante<sup>24,30,31</sup>. Esta es la única forma de saber si la posible falta de diferencia detectada se debe a un tamaño insuficiente de la muestra (equivalencia no es lo mismo que falta de diferencia) o a una falta de diferencia real, evitando de esta forma el gran problema de los “falsos negativos”. Por ello es importante, en el caso de resultados negativos, hacer mención, en la discusión, de la influencia del error beta y el poder estadístico en los resultados obtenidos<sup>10,29</sup>.

No resulta raro ver cómo suelen elegirse de manera arbitraria 20 o 30 pacientes, de forma que resulta imposible extraer cualquier tipo de conclusión (a no ser que haya grandes diferencias entre los tratamientos que se comparan). No hay que olvidar que con una muestra pequeña resulta difícil detectar diferencias reales, mientras que con una muestra suficientemente grande se puede detectar casi cualquier diferencia, por mínima que sea.

Las pruebas estadísticas que serán utilizadas en la sección “Resultados” deberían estar previamente indicadas en el apartado de “Material y métodos”, y tienen que ser las más adecuadas para el tipo de variables que se pretende analizar<sup>29</sup>.

Durante los últimos años han aparecido numerosas publicaciones<sup>28-31,34</sup> insistiendo en que resulta preferible expresar los resultados en forma de intervalos de confianza que como valor de *p* (no es lo mismo una *p* de 0,06 que una *p* de 0,4 y, sin embargo, en los 2 casos se habla de

diferencias no significativas), ya que dichos intervalos permiten obtener una idea más real de la magnitud de las diferencias observadas. A pesar de ello resulta muy difícil encontrar artículos en los que se siga este consejo. No obstante, la utilización de intervalos de confianza no debe eximir de realizar el cálculo del tamaño de la muestra<sup>35</sup>.

Es inaceptable expresar los resultados como tantos por ciento, extrayendo conclusiones de su comparación directa, sin realizar ningún tipo de análisis estadístico.

*Análisis de los resultados.* Deberían analizarse los pacientes asignados inicialmente a cada grupo de tratamiento, independientemente de que hayan completado o no el período de tratamiento y/o seguimiento. Esto es lo que se conoce como análisis por intención de tratar (*intention to treat*)<sup>1,18,19</sup>, y tiene una especial importancia en determinados tipos de ensayos, como por ejemplo los de antibióticos, ya que es la única forma de demostrar su eficacia real cuando sean utilizados en la práctica clínica diaria, donde muchas veces no es fácil la realización de cultivos y antibiogramas que permitan excluir a un paciente en el que aparecen resistencias.

El otro tipo de análisis, en el que sólo se incluyen los pacientes evaluables (p. ej., pacientes en los que el germen aislado es sensible al antibiótico a estudio, que han recibido un número mínimo de dosis del tratamiento, etc.) se conoce como análisis por tratamiento<sup>1,18</sup>. Éste puede ser realizado tras el anterior, aunque en determinados tipos de estudios es el más idóneo.

En el análisis de los resultados hay que considerar el número de pacientes que completaron el tratamiento y el motivo y posible repercusión de las pérdidas en cada grupo<sup>9,10</sup>. Resulta muy frecuente que se realicen análisis por subgrupos, de forma exploratoria y sin una hipótesis *a priori*<sup>28</sup>. Evidentemente, “cruzando” todas las variables es probable encontrar algún resultado significativo. Este tipo de comparaciones múltiples puede resultar útil para establecer nuevas hipótesis y planear ensayos adicionales encaminados a resolver una nueva pregunta, pero nunca para extraer conclusiones de los mismos.

*Conclusiones.* Las conclusiones deben estar basadas en los resultados obtenidos. Se entiende por validez interna la validez de los resultados para los pacientes del estudio<sup>36</sup>. Así mismo, hay que preguntarse si es posible generalizar los resultados a toda la población de pacientes, teniendo en cuenta los criterios de selección esta-

blecidos (validez externa). Es totalmente inaceptable establecer conclusiones derivadas de análisis realizados a subgrupos de manera exploratoria o *a priori*.

*Método de recogida de acontecimientos adversos (AA).* Según las recomendaciones sobre Buena Práctica Clínica de la Comunidad Europea y la Ley del Medicamento, se define como acontecimiento o evento adverso “cualquier experiencia no deseada que le ocurra a un sujeto durante un ensayo clínico, se considere o no, relacionada con los productos en investigación”<sup>11,37</sup>. Es preciso indicar los AA surgidos a lo largo del EC.

*Especificación si el ensayo se realizará según Buena Práctica Clínica (BPC).* La BPC son “normas según las cuales los ensayos clínicos son diseñados, realizados y comunicados de modo que se asegure que los datos son fiables y que se protegen los derechos, la integridad y la confidencialidad de los sujetos”<sup>36</sup>. En ellas se indican también las responsabilidades de los diferentes implicados en cada una de las fases de planificación y ejecución de un EC y obligan a la existencia de unos procedimientos preestablecidos (Procedimientos Normalizados de Trabajo) por escrito que se apliquen de forma sistemática durante todo el desarrollo del EC. En España, además, tenemos que tener en cuenta que la obligación de cumplir las normas de BPC viene explícita en la Ley del Medicamento.

*Aprobación por el Comité Ético de Investigación Clínica (CEIC).* Cada día son más las revistas que exigen para la publicación de un EC, la certificación de que tal EC ha sido aprobado por un CEIC (aunque ello no garantice de forma absoluta la ética del EC). En España, el Real Decreto que desarrolla la Ley del Medicamento<sup>38</sup> recoge la obligatoriedad de que se haga así y, además, detalla las funciones de los CEIC, que se podrían resumir en garantizar la protección de los sujetos participantes en el estudio.

*Consentimiento informado del paciente.* Se define el consentimiento informado como el proceso que garantiza que el sujeto ha expresado voluntariamente su intención de participar en el ensayo clínico, después de haber comprendido la información que se le ha dado acerca de los objetivos, beneficios, incomodidades y riesgos previstos, derechos y responsabilidades. Es imprescindible que el sujeto otorgue libremente su consentimiento informado antes de poder ser in-

cluido en un ensayo clínico y así debe constar en todo ensayo publicado.

*Check-list para evaluar la calidad de protocolos de ensayos clínicos (anexo II).*

En general, la mayoría de los apartados de este cuestionario se corresponden a los del anterior, sin embargo, por tratarse de protocolos de ensayo clínico no se hace referencia a alguno de ellos y es necesario hacer algunas consideraciones especiales. En el Real Decreto en el que se establecerán los requisitos para la realización de ensayos clínicos, se especifican los apartados básicos que debe incluir todo protocolo de ensayo clínico.

*Justificación del ensayo en función de la fase de desarrollo del fármaco.* Aunque este punto siempre debe ser considerado, es necesario que en todo protocolo de ensayo clínico se especifiquen las razones para llevarlo a cabo, y si los resultados que se prevé obtener justifican su realización.

*Acontecimientos adversos.* Deberían especificarse de forma clara los métodos de recogida, registro y notificación de acontecimientos adversos, valorándose de forma especial la existencia de una hoja de recogida de AA adecuada.

Debería indicarse el método de detección empleado, considerándose preferible la utilización conjunta de un *check-list* o cuestionario, la comunicación espontánea y la pregunta estándar<sup>27</sup>; los criterios de imputabilidad y el modo de notificación de los AA graves e inesperados a los CEIC y a las Autoridades Sanitarias.

*Medios materiales y humanos con los que cuenta el centro.* Cuando se evalúa un protocolo de ensayo clínico, aparte de todos los aspectos hasta ahora revisados, es necesario considerar algunos aspectos prácticos. Una de las misiones fundamentales de los Comités Éticos de Investigación Clínica es asegurarse si se ha realizado el cálculo del número de pacientes disponibles para poder reclutar el número calculado en el tiempo previsto. No realizar este cálculo es uno de los motivos fundamentales de que bastantes ensayos clínicos no lleguen a concluir, con las graves repercusiones éticas, científicas y económicas que ello conlleva.

También se asegurará que haya un responsable que supervise directamente el desarrollo del mismo y de que el centro cuenta con la infraestructura necesaria (p. ej., respecto a los aparatos) para poder realizar todo lo indicado en el protocolo.

ANEXO II  
CHECK-LIST PARA EVALUAR PROTOCOLOS DE ENSAYOS CLÍNICOS

			C A D I N A
<b>1.</b>	<b>¿Está justificada la realización del ensayo? (en función de la fase de desarrollo del fármaco):</b> .....		□ □ □ □ □
<b>2.</b>	<b>Definición del objetivo del estudio:</b> .....	+ ± - NA	□ □ □ □ □
	¿Existe un objetivo fundamental?.....	□ □ □ □	
	¿Se define una hipótesis <i>a priori</i> ?.....	□ □ □ □	C A D I N A
<b>3.</b>	<b>Criterios de selección de los pacientes:</b> .....	+ ± - NA	□ □ □ □ □
	¿Se especifican adecuadamente los criterios de inclusión? (edad, sexo, criterios diagnósticos, etc.) .....	□ □ □ □	
	¿Se especifican adecuadamente los criterios de exclusión?.....	□ □ □ □	C A D I N A
<b>4.</b>	<b>Definición del tratamiento experimental y control:</b> .....	+ ± - NA	□ □ □ □ □
	¿Es reproducible? (dosis, intervalo, duración, vía de administración, etc.) .....	□ □ □ □	
	¿Es el control adecuado? .....	□ □ □ □	
	¿Se especifica el tratamiento concomitante? .....	□ □ □ □	
	¿Sería necesario un período de lavado?.....	□ □ □ □	
	¿Se mide el cumplimiento y se especifica el método?.....	□ □ □ □	C A D I N A
<b>5.</b>	<b>Aleatorización:</b> .....	+ ± - NA	□ □ □ □ □
	¿Se asigna de forma aleatoria los sujetos del EC (aleatorización).....	□ □ □ □	
	¿Se especifica el método de aleatorización?.....	□ □ □ □	
	¿Es la aleatorización ciega?.....	□ □ □ □	
	¿Sería necesario estratificar? .....	□ □ □ □	
	¿Se necesitaría un periodo de preinclusión? .....	□ □ □ □	C A D I N A
<b>6.</b>	<b>Medición objetiva de los resultados (enmascaramiento):</b> .....	+ ± - NA	□ □ □ □ □
	¿Es el método de enmascaramiento adecuado?.....	□ □ □ □	
	¿Estudio no enmascarado?.....	□ □ □ □	
	¿Simple ciego?.....	□ □ □ □	
	¿Doble ciego o doble <i>dummy</i> ? .....	□ □ □ □	
	¿Se prevé la valoración de la eficacia del enmascaramiento? .....	□ □ □ □	C A D I N A
<b>7.</b>	<b>Variable principal de valoración (<i>end point</i>):</b> .....	+ ± - NA	□ □ □ □ □
	¿Es objetiva?.....	□ □ □ □	
	¿Es relevante desde el punto de vista clínico?.....	□ □ □ □	C A D I N A
<b>8.</b>	<b>Criterios de evaluación de la respuesta:</b> .....	+ ± - NA	□ □ □ □ □
	¿Es/son objetivo/s? .....	□ □ □ □	
	¿Es/son sensible/s y específico/s?.....	□ □ □ □	
	¿Es/son un/os método/s validado/s?.....	□ □ □ □	
	¿Se especifican los criterios de fracaso? .....	□ □ □ □	
	¿Es suficiente el periodo de seguimiento postratamiento? .....	□ □ □ □	C A D I N A
<b>9.</b>	<b>Pérdidas:</b> .....	+ ± - NA	□ □ □ □ □
	¿Se especifican las pérdidas prealeatorización? .....	□ □ □ □	
	¿Se especifican las pérdidas postaleatorización? .....	□ □ □ □	
	¿Abandonos ( <i>drop out</i> )? .....	□ □ □ □	
	¿Retiradas ( <i>withdrawal</i> )?.....	□ □ □ □	
	¿Pueden reemplazarse? .....	□ □ □ □	

(Continúa)

ANEXO II  
CHECK-LIST PARA EVALUAR PROTOCOLOS DE ENSAYOS CLÍNICOS (Continuación)

	+ ± - NA	C A D I NA
<b>10. Diseño estadístico:</b> .....	+ ± - NA	□ □ □ □ □
¿Se calcula <i>a priori</i> el número de pacientes? .....	□ □ □ □	
¿Se tiene en cuenta el error alfa? .....	□ □ □ □	
¿Se tiene en cuenta el poder estadístico? .....	□ □ □ □	
¿Se tienen en cuenta las diferencias clínicamente relevantes? .....	□ □ □ □	
¿Se tienen en cuenta los abandonos? .....	□ □ □ □	
¿Se especifican los test estadísticos y son adecuados en función de las variables que hay que analizar? .....	□ □ □ □	C A D I NA
<b>11. Acontecimientos adversos (AA)</b> .....	+ ± - NA	□ □ □ □ □
¿Existe una hoja de recogida de AA? .....	□ □ □ □	
¿Se especifica el método de notificación de AA? .....	□ □ □ □	
¿Se especifica el método de imputabilidad? .....	□ □ □ □	C A D I NA
<b>12. ¿Se especifica si el ensayo se hará según BPC?</b> .....		□ □ □ □ □
<b>13. ¿Se ha solicitado la aprobación del protocolo por un Comité ético de investigación clínica?</b> .....		□ □ □ □ □
<b>14. ¿Se solicita el consentimiento informado a los sujetos participantes en el EC?</b> .....		□ □ □ □ □
<b>15. Medios materiales y humanos para realizar el ensayo:</b> .....	+ ± - NA	□ □ □ □ □
¿Es suficiente el número de pacientes disponibles? .....	□ □ □ □	
¿Son suficientes los recursos materiales? .....	□ □ □ □	
¿Existe un responsable directo de la supervisión del ensayo? .....	□ □ □ □	

+ = adecuado; ± = dudoso; - = inadecuado; NA = no aplicable; C = correcto; A = aceptable; D = dudoso; I = incorrecto; NA = no aplicable.

## Discusión

Durante los últimos años han sido publicados distintos tipos de cuestionarios o *check-lists* para evaluar los EC publicados<sup>8-15</sup>. Aunque los objetivos generales de todos ellos son los mismos, la extensión, formato y método de evaluación varía considerablemente de unos a otros. Existen dos tipos fundamentales de *check-lists*: los cuantitativos y los cualitativos<sup>1</sup>. En los primeros, a través de escalas numéricas, se asigna una puntuación a cada uno de los apartados, dependiendo de su importancia<sup>10,11</sup>, de forma que al final se da una puntuación global. En los segundos, por el contrario, la valoración no está basada en ese tipo de escalas y tiene una mayor importancia la evaluación de cada uno de los puntos<sup>8,12,13</sup>.

Ambos métodos tienen ventajas e inconvenientes. Los cuestionarios cuantitativos tienen las ventajas de una (teóricamente) mayor objetividad y de permitir una evaluación global del ensayo (p. ej., se han utilizado, para comprobar el aumento de la

calidad de los ensayos clínicos a lo largo del tiempo<sup>39</sup>); sin embargo, presentan serios problemas conceptuales: ¿se puede ser rígido con una escala de medida flexible? Establecer una determinada puntuación no deja de ser una forma de objetivar algo que a menudo es necesariamente subjetivo: ¿puede darse el mismo valor a varios apartados “menores” que un solo punto “mayor”? Sería demasiado simplista clasificar los ensayos en correctos e incorrectos basándose en una puntuación general. Realmente la calidad de casi todos los estudios es intermedia: aportan información valiosa, pero pueden ser mejorados en determinados aspectos<sup>9</sup>. Por fin, lo que consideramos su mayor inconveniente, un mismo apartado puede tener diferente valor dependiendo del tipo de estudio (p. ej., cuanto más subjetiva sea la variable principal de valoración mayor valor tendrá el enmascaramiento).

Por otro lado, los cuestionarios cualitativos permiten un análisis más matizado de cada uno de los apartados del ensayo, dando una idea más

clara de cuáles son los aspectos problemáticos. Su inconveniente es que con frecuencia constan de pocos apartados y son demasiado generales, llegando a perder su carácter de "herramienta". Exagerando, si se propone un *check-list* con dos apartados como son: diseño adecuado y aspectos éticos correctos, se englobará todo lo que debe evaluarse, pero su utilidad práctica será escasa.

No se pretende que los cuestionarios que aquí se presentan sean utilizados como unos instrumentos rígidos de medida (todo *check-list* es subjetivo y, por tanto, criticable), sino que sirvan de guía para repasar de forma sistemática los principales puntos que deben ser considerados cuando, a la vista de un determinado ensayo clínico, se quiera analizar la forma en que ha sido realizado, y así formarnos una idea de la solidez de sus conclusiones. Por todo ello, se ha elegido un tipo de cuestionario cualitativo, no excesivamente extenso, pero que analice de forma ordenada los aspectos que se consideran esenciales. Ambos cuestionarios constan de 15 apartados generales que deben ser evaluados individualmente utilizando las preguntas que aparecen en cada uno de los mismos.

Aunque no se han incluido en los *check-lists*, porque en estos momentos prácticamente ningún ensayo clínico los contempla, todo parece apuntar que durante los próximos años será necesario tener en cuenta dos aspectos adicionales: *a)* un apartado sobre las limitaciones del ensayo clínico, tales como posibles sesgos, validez de los resultados, etc., y grado en que las mismas pueden influir en las conclusiones. En determinadas revistas científicas ya se exige dicho apartado, y *b)* el impacto económico de los nuevos fármacos. Así, se ha desarrollado una nueva disciplina, la farmacoeconomía, definida como la descripción y el análisis de los costes del tratamiento farmacológico a los sistemas de salud y a la sociedad. Aunque se realicen estudios específicamente diseñados para evaluar el coste-beneficio, coste-efectividad, etc., es muy probable que la evaluación económica sea también un objetivo secundario de ensayos clínicos diseñados primeramente para evaluar eficacia y seguridad<sup>40</sup>. Este tipo de análisis económico, para mantener su credibilidad, debe tener el mismo rigor metodológico que cualquier ensayo clínico ya que, por naturaleza, es especialmente vulnerable a los sesgos. En este sentido, algunos países, como Australia, han propuesto directrices especiales para el desarrollo de estos estudios<sup>41</sup>.

Un problema ya apuntado por algunos autores es que en muchas ocasiones lo publicado es me-

nos de lo realizado<sup>10</sup>. Desgraciadamente, por motivos de espacio es necesario acortar alguna de las secciones más importantes (habitualmente la de "Material y métodos"). Tampoco suele incluirse un apartado esencial para valorar a fondo la calidad del ensayo: el cuaderno de recogida de datos (CRD). Aunque es fácil de comprender, no deja de ser sorprendente que a menudo se dé tan poca importancia a uno de los aspectos más cruciales de la investigación: la reproducibilidad. En este punto la investigación básica saca una importante ventaja a la clínica, y es quizás el motivo de que, en general, la fiabilidad de sus resultados sea superior. Para lograr ese objetivo se debería escribir el apartado de "Material y métodos" de una manera más detallada, y esa es una de las tareas en la que todos debemos colaborar. La unificación de criterios que puede obtenerse utilizando cuestionarios como los que aquí se presentan puede ser una forma de contribuir a mejorar la interpretación de los datos de los ensayos clínicos, y en última instancia a lograr una mejor utilización de los medicamentos.

## BIBLIOGRAFÍA

1. Spilker B. Guide to clinical trials. Nueva York: Raven Press, 1991.
2. Klimt CR. The conduct and principles of randomized clinical trials. *Controlled Clin Trials* 1981; 1: 283-293.
3. Higgins MW. Conducting clinical trials. *Am Rev Resp Dis* 1980; 122: 3-9.
4. Green SB. Randomized clinical trials: design and analysis. *Semin Oncol* 1981; 8: 417-423.
5. Spilker B. Practical considerations in planning and conducting clinical trials with investigational or marketed drugs. *Clin Neuropharmacol* 1983; 6: 325-347.
6. Der Simonian R, Charette LJ, Mc Peek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1.332-1.337.
7. Gardner MJ, Bond J. An exploratory study of statistical assessment of paper published in the *British Medical Journal*. *Br Med J* 1990; 263: 1.355-1.357.
8. Grant A. Reposting controlled trials. *Br J Obstet Gynaecol* 1989; 96: 397-400.
9. Fowkes FGR, Fulton PM. Critical appraisal of published research: introductory guidelines. *Br Med J* 1991; 302: 1.136-1.140.
10. Chalmers TC, Smith H Jr, Balckburn B, Silverman B et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2: 31-49.
11. Evans RW, Pollock AV. A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infections. *Br J Surg* 1985; 72: 256-260.

12. Gardner MJ, Machin D, Campbell MJ. Use of check lists in assessing the statistical content of medical studies. *Br Med J* 1986; 292: 810-812.
13. Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Assoc J* 1964; 90: 565-569.
14. Weintraub M. How to critically assess clinical drug trials. *Drug Ther* 1982; 12: 131-148.
15. Dixon JS, Smith A, Evans SJV. Reporting clinical trials. *Br J Rheumatol* 1983; 22 Supl: 74-78.
16. British Medical Journal. Guidelines for writing papers. *Br Med J* 1988; 296: 48-50.
17. Ley del Medicamento. Ministerio de Sanidad y Consumo. BOE 306, 22 de diciembre de 1990.
18. Grupo de Trabajo sobre Terminología. Coordinadora: Dra. Inés Galende. Glosario de términos de ensayos clínicos. En: García Alonso F, Bakke OM, editores. Metodología del ensayo clínico. Barcelona: Doyma, 1991; 113-116.
19. Galende I, Tristán C. Problemas prácticos en un ensayo clínico (II). En: García Alonso F, Bakke OM, editores. Metodología del ensayo clínico. Barcelona: Doyma, 1991; 21-30.
20. Cramer JA, Mattson RH, Prevey ML et al. How often is medication taken as prescribed? A novel assessment technique. *JAMA* 1989; 261: 3,273-3,277.
21. Bailar JC III, Louis TA, Lavori PW, Polansky M. Studies without internal controls. *N Engl J Med* 1984; 311: 156-162.
22. Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-153.
23. Altman DG. Randomisation. Essential for reducing bias. *Br Med J* 1991; 302: 1,481-1,482.
24. Pocock SJ. Current issues in the design and interpretation of clinical trials. *Br Med J* 1985; 290: 39-42.
25. Meinert CI. Clinical trials: design, conduct and analysis. Oxford: Oxford University Press, 1986.
26. Warlow C. Organise a multicentre trial. *Br Med J* 1990; 300: 180-183.
27. De Abajo M, Serrano-Castro MA. Problemas prácticos en un ensayo clínico (I). En: García Alonso F, Bakke OM, editores. Metodología del ensayo clínico. Barcelona: Doyma, 1991; 11-20.
28. 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: 426-432.
29. Bailar III JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med* 1988; 108: 266-273.
30. Carné X, Moreno V, Porta M, Velilla E. El cálculo del número de pacientes necesarios en la planificación de un estudio clínico. *Med Clin (Barc)* 1988; 92: 72-77.
31. Porta M, Moreno V, Sanz F, Carné X, Velilla E. Una cuestión de poder. *Med Clin (Barc)* 1989; 92: 223-228.
32. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J* 1986; 292: 746-750.
33. Simon R. Confidence intervals for reporting results of clinical trials. *Ann Intern Med* 1986; 105: 429-435.
34. Plasencia A, Porta M. La calidad de la información clínica (II): significación estadística. *Med Clin (Barc)* 1988; 90: 122-126.
35. Daly LE. Confidence intervals and sample sizes: don't throw out all your old sample size tables. *Br Med J* 1991; 302: 333-336.
36. Porta M, Alvarez-Dardet C, Bolúmar F, Plasencia A et al. La calidad de la información clínica (I): validez. *Med Clin (Barc)* 1987; 89: 741-747.
37. CPMP Working Party on Efficacy of Medicinal Products. EEC note for guidance: Good Clinical Practice for trials on medicinal products in the european community. *Pharmacol Toxicol* 1990; 67: 361-372.
38. Proyecto de Real Decreto, que desarrolla el Título Tercero ("De los ensayos clínicos"), de la Ley 25/1990, de 20 de diciembre, del Medicamento.
39. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clin Trials* 1990; 11: 339-352.
40. Bootman JL, Larson LN, McGhan WF, Townsend RJ. Pharmacoeconomic research and clinical trials: concepts and issues. *Drug Intell Clin Phar* 1989; 23: 693-697.
41. Commonwealth of Australia. Guidelines of the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including submissions involving economic analysis. Canberra (ACT): Department of Health, Housing and Community Services, 1990.



# Cómo mejorar la calidad de los ensayos clínicos

Inés Galende Domínguez<sup>a</sup>, José A. Sacristán<sup>b</sup> y Javier Soto<sup>c</sup>

<sup>a</sup>Dirección General de Farmacia y Productos Sanitarios. Ministerio de Sanidad y Consumo. Madrid.

<sup>b</sup>Departamento de Investigación Clínica, Lilly S.A.

<sup>c</sup>Unidad 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 validez<sup>1</sup>, hecho que es agravado por la no publicación de los EC con resultados negativos<sup>2</sup>.

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 esperados<sup>3</sup>.

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 grupo<sup>4</sup> en el diseño, conducción y publicación de los EC al aplicar una lista guía propuesta previamente<sup>5</sup>. 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<sup>6,7</sup>. 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 respuesta<sup>8</sup>.

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 tratamiento<sup>9</sup>.

Las características que debe cumplir un buen método de aleatorización<sup>10</sup> 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 matemático<sup>11</sup>. 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 método<sup>11</sup> 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 reclutamiento<sup>12</sup>.

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<sup>10,13</sup>.

**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 50<sup>13</sup>.

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 global<sup>11</sup>.

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 Zelen<sup>14</sup> 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 inaceptables<sup>10</sup>. 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<sup>10,13</sup>.

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 clínicamente<sup>11</sup>.

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 recomendaciones<sup>10</sup>:

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 tratamientos<sup>9</sup>.

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 administrado<sup>9</sup>. 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 control<sup>9</sup>.

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 grave<sup>15</sup>. 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 ensayo<sup>16</sup>.

### **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 situación<sup>12</sup>.

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 tratamiento<sup>12</sup>. 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 1978<sup>17</sup> 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 resultados<sup>18</sup> 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 1978<sup>19</sup> principalmente los de no maleficencia y autonomía que son los más afectados en los casos descritos<sup>20</sup>.

Con el objeto de mejorar la calidad de los ensayos clínicos, algunos autores<sup>21</sup> 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 clínico<sup>3</sup>.

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.
- Gotsche 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.

---

# La buena práctica clínica y la caja de Pandora (y III). La fiabilidad de los datos

---

Carles Vallvé

Servei de Farmacologia Clínica. Institut Català de Farmacologia. Universitat Autònoma de Barcelona.  
Ciutat Sanitària i Universitària Vall d'Hebron. Barcelona.

La revisión del protocolo por un comité ético y la obtención del consentimiento informado son los procedimientos previstos en la Unión Europea<sup>1</sup> para proteger los derechos de los pacientes incluidos en un ensayo clínico. Para proteger a la población a la que posteriormente se aplicarán los resultados del ensayo, será necesario considerar la validez científica del protocolo y se deberán tomar medidas que conduzcan a demostrar la fiabilidad de los datos que se obtengan en el mismo. Estas medidas consisten en la adopción de procedimientos de "garantía de la calidad" por parte del promotor del ensayo y en la realización de inspecciones por parte de las autoridades sanitarias.

El ensayo clínico, como la misma práctica de la medicina, es un proceso dinámico. En el ámbito de la medicina, el "control de la calidad" sólo es posible en el laboratorio clínico: se obtiene una alícuota de una muestra biológica y la prueba se puede repetir en otro laboratorio si las condiciones de conservación han sido adecuadas. Pero no es posible repetir la exposición a un medicamento para comprobar que ha determinado mejoría de la enfermedad. Tampoco es posible repetir el proceso diagnóstico una vez ha curado la enfermedad. Para comprobar la validez de los datos que constan en un cuaderno de recogida de datos, es necesario recurrir a medidas indirectas: por ejemplo, cotejarlos con los que constan en la historia clínica del paciente, lo que presupone admitir que en ésta se reproduce exactamente la realidad clínica. Por este motivo no se habla de "control de la calidad", sino de "garantía de la calidad".

La garantía de la calidad implica, por parte del promotor, la disponibilidad de procedimientos normalizados de trabajo (PNT) y la realización de auditorías. Es evidente que el responsable de la aplicación de los procedimientos de garantía de

calidad no puede estar directamente implicado en la realización del ensayo.

¿Por qué es necesaria la garantía de la calidad? Simplemente, para descartar la posibilidad de errores y para descartar la posibilidad de fraude.

## Errores y fraudes en los ensayos clínicos

---

Es tan difícil evitar errores metodológicos en el ensayo clínico como errores diagnósticos en la práctica médica. Los errores metodológicos no pueden subsanarse mediante la aplicación de procedimientos de garantía de la calidad. Pero estos procedimientos pueden descubrir los llamados "errores humanos".

Si se introducen por duplicado los datos de un cuaderno de registro en una base de datos informatizada, y la introducción es efectuada por dos personas diferentes, es muy difícil que se produzcan los mismos errores, de forma que al confrontar ambas bases de datos se podrá comprobar que las divergencias se deben a errores de lectura o de transcripción cometidos por los responsables de la introducción de los datos. Lo sorprendente es que si se toma una muestra de datos al azar del documento depurado y se coteja con el documento original, no es imposible que se descubran nuevos errores, cometidos bien en la lectura, bien en el registro de los mismos. Este procedimiento permite mantener el porcentaje de errores a un nivel muy bajo, es decir, que, con una probabilidad elevada, no influirán cuantitativamente en los resultados del ensayo.

Como ha indicado Dal-Ré, aunque en España no se han hecho públicos casos de fraude en ensayos clínicos, es lógico pensar que ocurren<sup>2</sup>. De hecho, en EE.UU., Australia y el Reino Unido, en los últimos 15 años se han descrito multitud de casos de fraude en ensayos clínicos<sup>3</sup>. Esta discrepancia se debe probablemente a que en España, como en otros países, no existen procedimientos que determinen la conducta del pro-

motor del ensayo en caso de fraude o de sospecha de fraude. Los procedimientos de garantía de la calidad pueden permitir descubrir precozmente la existencia de fraude y, en caso dado, adoptar las medidas correctoras necesarias. Una actitud *estricta* en la monitorización de ensayos clínicos es quizás el mejor medio para evitar el fraude o, en ocasiones, para corregir errores: al revisar los cuadernos de registro de casos de un ensayo controlado se comprobó que según edad, sexo e iniciales, 2 casos debían de corresponder a un mismo paciente; dado que los restantes datos eran prácticamente idénticos (con la excepción de las fechas de registro), surgió la sospecha de que el paciente había sido incluido de manera fraudulenta en el ensayo; en una entrevista con el investigador se comprobó que se trataba de una paciente que 5 días después de haber sido incluida en el ensayo, desarrolló lesiones dermatológicas de características idénticas a las que motivaron su inclusión en el ensayo. Esta situación no estaba prevista en el protocolo, por lo que el investigador consideró que se trataba de un nuevo proceso patológico y que era adecuado incluir a la paciente de nuevo en el ensayo. Si se hubiera tratado realmente de un caso de fraude, hubiera sido necesario, bien: *a*) interrumpir el ensayo, excluir al investigador clínico del desarrollo del producto y comunicar esta decisión al correspondiente comité ético, lo que puede resultar contrario (al menos aparentemente) a los intereses del promotor, o bien *b*) excluir al paciente, comprobar la validez de los datos de todos los casos restantes y monitorizar de forma estricta el ulterior desarrollo del ensayo.

Como es lógico, en estos casos no es posible ofrecer normas rígidas de conducta.

### **Los procedimientos normalizados de trabajo**

Se ha mencionado que los PNT tienen en investigación clínica el mismo valor que el Código de Circulación en la carretera. Si no existe una normativa previa, no es posible decidir imparcialmente si se ha cometido o no una infracción.

El Real Decreto 561/1993 exige que el promotor disponga por lo menos de los siguientes PNT<sup>4</sup>:

*a*) Identificación y calificación del investigador principal y sus colaboradores; *b*) archivo de la documentación esencial del ensayo clínico; *c*) monitorización, incluyendo verificación de los datos originales; *d*) suministro de la medicación en estudio; *e*) notificación de acontecimientos adversos graves e inesperados, y *f*) comprobación de que el sujeto ha otorgado su consentimiento informado para participar en el ensayo.

Posiblemente la adecuada realización de un ensayo clínico exige también disponer de normas adicionales, como la destinada a uniformizar el contenido y la estructura de los informes de ensayo clínico<sup>5</sup>, que actualmente se está elaborando en el contexto de la *International Conference on Harmonization*, cuyo objetivo es estandarizar las exigencias de las autoridades sanitarias europeas, japonesas y estadounidenses en materia de registro farmacéutico.

Pero resulta sorprendente el gran número de PNT que han sido elaborados por las unidades de garantía de calidad de los grandes promotores de ensayos clínicos: no es excepcional que un laboratorio disponga de más de 40 PNT, en los que se detallan en forma telegráfica hasta las más simples operaciones que deben realizarse en el curso de un ensayo clínico como, por ejemplo, la forma de ordenar la impresión de un determinado número de cuadernos de recogida de datos ("revisar una prueba de imprenta antes de ordenar la impresión de todos los ejemplares requeridos"). Existen pocas dudas de que los PNT son un excelente instrumento para uniformizar actividades dentro de un Departamento de Investigación y Desarrollo, pero no es extraño que determinadas disposiciones tengan poco que ver con la calidad del ensayo clínico.

El *recuento de pastillas* es un buen ejemplo de la afirmación anterior. En 3 estudios publicados en 1989 se demostró la escasa o nula utilidad de esta medida. Pullar et al<sup>6</sup>, utilizando pequeñas cantidades de fenobarbital añadido a la formulación galénica, comprobaron que el recuento de muestras exagera ampliamente el grado de cumplimiento. Rudd et al<sup>7</sup> analizaron la variancia del recuento de muestras, concluyendo que evidenciaba la existencia de fraude por parte del paciente. Cramer et al<sup>8</sup> lo compararon con las concentraciones plasmáticas del fármaco y la supervisión continua de la dosificación, y dedujeron que el recuento no demuestra si el paciente toma una dosis mayor, menor o igual a la prescrita. No obstante, en la actualidad, esta es una actividad que se encuentra prevista en casi todos los protocolos de ensayo clínico.

La *documentación del investigador (Investigator's Study File)* debe estar a disposición de las autoridades sanitarias para que, en caso de inspección, éstas puedan cerciorarse de que conoce el protocolo del ensayo y las recomendaciones que se deben seguir para desarrollarlo correctamente. Resulta sorprendente que un PNT disponga que el monitor del ensayo o el adjunto de investigación clínica (CRA) deba cerciorarse de que el investigador archiva correctamente las co-

pías de los escritos que *él mismo* ha dirigido al promotor del ensayo. En realidad, es norma habitual que el promotor prepare en un archivero adecuado la documentación que el investigador debe tener disponible en caso de inspección, lo que podría ser juzgado como un intento de desinformar a las autoridades sanitarias.

Lo tratado en las visitas de monitorización debe quedar reflejado en un informe del monitor o del adjunto de investigación clínica (CRA): estos informes tienen el valor de mantener al promotor del ensayo al tanto del desarrollo del mismo, permiten comprobar si se presta al ensayo toda la atención requerida y si se cumplen las diferentes decisiones que ha sido preciso adoptar en su desarrollo. La tendencia de que en el centro donde se realiza el ensayo se disponga también de un *cuaderno de monitorización (monitoring log)*, en el que quede constancia de la fecha de las visitas de monitorización, junto con las firmas del monitor o del CRA y del investigador clínico, parece un procedimiento burocrático que sólo evidencia la falta de confianza del promotor en sus colaboradores, y que difícilmente puede contribuir a incrementar la calidad de un ensayo.

### **Auditorías e inspecciones**

Varios promotores de ensayos clínicos disponen en España de unidades de garantía de calidad que realizan auditorías de ensayos clínicos; se ha calculado que el coste de la implantación de las normas de buena práctica clínica supone un incremento del 8 % en el presupuesto de I + D<sup>9</sup>. Otros promotores efectúan auditorías en el marco de programas internacionales de garantía de la calidad<sup>10</sup>. Como es lógico, la eficiencia de estas medidas depende del respaldo que reciban por parte de las autoridades sanitarias. La Dirección General de Farmacia y Productos Sanitarios inició en 1992 un programa de inspecciones de ensayos clínicos que, lamentablemente, no pasó de su fase piloto: sólo se efectuaron varias inspecciones de ensayos de bioequivalencia y una inspección a un investigador clínico por sospecha de fraude<sup>11</sup>.

La falta de una vigilancia adecuada puede resultar en que, en las carreteras de la investigación clínica, pronto veamos a promotores e investigadores clínicos que no muestran el debido respeto a las señales de tráfico. No sería extraño que en estas circunstancias, alguien, parodiando a la *General Accounting Office* de 1976, llegara a afirmar que “el control estatal del ensayo con nuevos medicamentos no protege adecuadamente a los sujetos humanos incluidos en los mismos ni al público en general”<sup>12</sup>.

### **La calidad científica de un ensayo clínico**

Que un ensayo se realice de acuerdo con las normas de buena práctica clínica, así como que el promotor disponga de procedimientos normalizados de trabajo, no proporciona ninguna información acerca de la calidad científica del protocolo del ensayo clínico. Tampoco ayuda al comité ético a evaluar la eficiencia científica del protocolo, es decir, a evaluar “la posibilidad de alcanzar conclusiones válidas, con la menor exposición posible del sujeto”, según definición del Real Decreto.

Para evaluar la calidad de un ensayo clínico se han propuesto diferentes métodos: *a)* un esquema de revisión de protocolos<sup>13</sup>; *b)* una lista de palabras clave<sup>14</sup>; *c)* una lista de preguntas<sup>15</sup>, y *d)* dos procedimientos ponderados de evaluación<sup>16,17</sup>.

Es importante tener en cuenta que la *Food and Drug Administration* acepta que un ensayo clínico aporta “evidencia sustancial” acerca de la eficacia de un producto sólo cuando es posible calificarlo como “adecuado y bien controlado”<sup>18</sup>, es decir, cuando: *a)* los objetivos están claramente expuestos; *b)* el diseño experimental permite una comparación válida con el grupo de control; *c)* el procedimiento de selección de pacientes demuestra adecuadamente que los sujetos padecen la enfermedad que se estudia; *d)* el método de asignación de los pacientes a los tratamientos asegura la formación de grupos comparables; *e)* se adoptan las medidas necesarias para reducir al máximo los sesgos de los sujetos y de los observadores y en la valoración de resultados; *f)* la respuesta de los pacientes se registra mediante métodos adecuados y bien definidos, y *g)* los resultados del ensayo se analizan de forma adecuada.

En todo caso, estas exigencias (y también los procedimientos normalizados de trabajo, que aseguran que se han cumplido en forma adecuada) se basan en la teoría de que el conocimiento científico se obtiene a partir de la observación, que permite extraer enunciados singulares. Cuando los enunciados singulares se repiten suficientes veces, se infieren enunciados universales; es la *teoría inductivista* del conocimiento científico<sup>19</sup>. Si se acepta, como propone Feyerabend, que todo método implica una simplificación indebida, si aceptáramos que no existen reglas para definir *a priori* el conocimiento científico y que la idea de un método fijo se apoya en una concepción ingenua del hombre y de su entorno social, deberíamos preguntarnos si el método del ensayo clínico no es interpretado y utilizado como un dogma y que, si se sigue, es sólo por razones de comodidad<sup>20</sup>.

Pero, sin ir tan lejos como Feyerabend, tal vez sería necesario preguntarse, por los motivos expuestos, si algunos aspectos de la garantía de la calidad en el ensayo clínico, tal como se practica actualmente, no constituyen más un impedimento que un estímulo real para el progreso del desarrollo farmacológico.

### Agradecimiento

El autor agradece al Prof. J.R. Laporte la cuidadosa revisión y los valiosos comentarios aportados a los 3 artículos de "La buena práctica clínica y la caja de Pandora".

### BIBLIOGRAFÍA

1. CPMP Working Party on Efficacy of Medicinal Products. Good Clinical Practice for trials on medicinal products in the European Community. *Pharmacol Toxicol* 1990; 67: 361-372.
2. Dal-Ré R. Comités Éticos de Investigación Clínica: algo más que el cambio de nombre. *Med Clin (Barc)* 1995; 105: 580-582.
3. Lock S. Research misconduct: a résumé of recent events. En: Lock S, Wells F, editores. *Fraud and misconduct in medical research*. Londres: BMJ Publishing, 1993; 5-24.
4. Ministerio de Sanidad y Consumo. Real Decreto 561/1993, de 16 de abril, por el que se establecen los requisitos para la realización de ensayos clínicos con medicamentos. Madrid: BOE 1993; 114: 14.346-14.364.
5. The European Agency for the Evaluation of Medicinal Products. Note of guidance on structure and content of clinical study reports. Draft 13. 28 de marzo, 1995.
6. Pullar T, Kumar S, Tindall H, Feely M. Time to stop counting the tablets? *Clin Pharmacol Ther* 1989; 46: 163-168.
7. Rudd P, Byyny RL, Zachary V, LoVerde ME, Titus C, Mitchell WD, Marshall G. The natural story of medication compliance in a drug trial: limitations of pill counts. *Clin Pharmacol Ther* 1989; 46: 169-176.
8. Cramer JA, Mattson RH, Prevey ML, Scheyer RD, Oulette VL. How often is medication taken as prescribed? A novel assessment technique. *JAMA* 1989; 261: 3.273-3.277.
9. Vilumara A. Coste económico en el seguimiento de las BLP y PBC. En: *Calidad de vida y farmacoeconomía*. Barcelona: Esame, 1996: 40-44.
10. Dal-Ré R, García-Losa M. Auditorías en la investigación clínica con fármacos. *Med Clin (Barc)* 1992; 98: 311-313.
11. Avendaño C. Inspecciones de ensayos clínicos. I Congreso Nacional de Medicina Farmacéutica. Madrid: AMIFE, 1993.
12. General Accounting Office. Report to Congress. Federal control of new drug testing is not adequately protecting human test subjects and the public. 15 de julio de 1976.
13. Perrier CV, Assai JP, Donath A, Roux JL. Directives pour la rédaction d'un protocole de recherche. *Méd Hyg* 1975; 33: 461-462.
14. Clinical Trials Unit. Department of Pharmacology and Therapeutics. London Hospital Medical College. Aide-mémoire for preparing clinical trial protocols. *Br Med J* 1977; 1: 1.323-1.324.
15. Spriet A, Simon P. *Methodology of clinical drug trials*. Basilea: Karger, 1985.
16. Vallvé C. Buena Práctica Clínica. Recomendaciones internacionales en investigación terapéutica. Madrid: Farmaindustria, 1990.
17. Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2: 31-49.
18. Food and Drug Administration. Draft guideline for the format and content of the clinical data section of an application. (Docket no 85D-0467). Enero de 1986.
19. Chalmers AF. ¿Qué es esa cosa llamada ciencia? Madrid: Siglo XXI, 1989.
20. Feyerabend PK. *Adiós a la razón*. Madrid: Tecnos, 1987.

---

# *En busca de la evidencia perdida.* Propuestas para incrementar la transparencia de los ensayos clínicos

---

Gerard Urrutia<sup>a,b</sup>, Xavier Bonfill<sup>a</sup> y Joan Martí<sup>a</sup>

<sup>a</sup>Centro Cochrane Español (Fundación Parc Taulí). <sup>b</sup>Secretario del Comité Ético de Investigación Clínica. Corporació Sanitària Parc Taulí. Sabadell. Barcelona.

## **La transparencia de los ensayos clínicos y sus problemas**

---

Está ampliamente aceptado que los ensayos clínicos constituyen la mejor herramienta de la que disponemos (estándar de oro) para evaluar la eficacia de las intervenciones sanitarias (sobre todo, aunque no exclusivamente, terapéuticas), proporcionando el mayor grado posible de evidencia científica al respecto<sup>1-5</sup>. Esto es así ya que su diseño –comparación entre grupos, con asignación aleatoria y evaluación objetiva, como rasgos fundamentales– permite minimizar la posibilidad de sesgos o errores sistemáticos asegurando, de esta manera, la validez interna o fiabilidad de los resultados obtenidos. Los ensayos clínicos contribuyen, por tanto, a incrementar el acervo de conocimiento científico sólido y válido que debe sustentar la práctica clínica de manera especialmente significativa.

No obstante, los resultados de los ensayos clínicos deben ser convenientemente difundidos proporcionándose, además, información lo suficientemente detallada sobre sus características (diseño y ejecución) y los resultados alcanzados, de manera que los lectores puedan formarse un juicio propio razonado y bien fundamentado sobre su validez. Por contra, si la existencia de dichos ensayos clínicos, o sus resultados, no es públicamente conocida, mediante su oportuna difusión (principalmente mediante artículos en revistas) o bien si los informes de sus resultados son de calidad deficiente, difícilmente podrán ser de alguna utilidad.

Uno de los principales problemas de la no publicación de los resultados de un ensayo clínico es el denominado *sesgo de publicación*. Varios estudios dirigidos de forma específica a valorar las posibles causas de este sesgo han puesto en

evidencia cómo la magnitud y la dirección de los resultados hallados por los ensayos clínicos determinan en gran manera su probabilidad de ser publicados<sup>6-19</sup>. Así, los ensayos clínicos que concluyen con resultados “positivos” (es decir, que encuentran una diferencia estadísticamente significativa entre las diversas pautas comparadas) tienen una mayor probabilidad de ser comunicados y publicados frente a los ensayos clínicos que concluyen con resultados “negativos”. Además, cada uno de estos estudios genera un mayor número de publicaciones, con menor grado de retraso y en revistas de mayor impacto. Esta tendencia se asocia con la importancia que el mismo investigador atribuye a los resultados obtenidos minusvalorando la relevancia de unos resultados “negativos”, a la actitud restrictiva de editores y revisores de revistas que, ante las limitaciones de espacio, seleccionan sólo los ensayos clínicos que consideran de mayor impacto, generalmente los más prometedores<sup>20-23</sup> y al interés de la industria farmacéutica, promotora de la mayoría de ensayos clínicos, para no divulgar resultados no claramente favorables.

La calidad de los informes y publicaciones de los ensayos clínicos no es, en absoluto, una cuestión de importancia secundaria. De esta forma, resulta imprescindible que el lector pueda valorar si la calidad de la información que éstos contienen es razonable o, dicho de otro modo, que existe validez interna –definida como la garantía de que el diseño del estudio, su ejecución, análisis y presentación de los resultados minimizan o evitan los sesgos en las comparaciones entre las intervenciones que están siendo evaluadas–. Esta definición sobre la calidad de los informes se basa, pues, en la evaluación de diversos factores específicos que sabemos pueden influir en los resultados de una intervención. Se hace necesario, pues, consensuar por parte de los equipos editoriales de las revistas biomédicas

dicas, la administración sanitaria y los organismos científicos unos criterios mínimos que deberían cumplirse en todo caso para la publicación y la posterior evaluación crítica de los ensayos clínicos.

En resumen, el conocimiento que no ha sido divulgado no puede suponer ninguna contribución al dejar de repercutir sobre la información disponible para la comunidad científica y, por tanto, no contribuirá en modo alguno a una más apropiada toma de decisiones clínicas ni a la formulación de una política sanitaria más acertada. Sólo mediante la divulgación de los resultados de la investigación científica será posible que la ciencia logre alcanzar su objetivo principal, *la mejora del bienestar del ser humano*, ya que la comunicación de los ensayos clínicos de calidad, realizados de forma adecuada, constituye la base para una atención sanitaria más segura y efectiva. Bajo esta misma asunción, los autores proponen en este artículo una serie de medidas, algunas de ellas enunciadas o incluso puestas en práctica por otros autores u organismos, para incrementar la transparencia en la realización y publicación de los resultados de los ensayos clínicos.

### **Medidas propuestas para incrementar la transparencia en la realización y publicación de los resultados de los ensayos clínicos**

#### *Registro prospectivo y accesible de los ensayos clínicos autorizados por la autoridad sanitaria*

Varios autores han propuesto el establecimiento de registros prospectivos de ensayos clínicos como prerrequisito para su autorización y aprobación por parte de la autoridad sanitaria y por los comités éticos locales o regionales<sup>24-26</sup>. España ha sido uno de los países pioneros en el establecimiento de esta medida por parte de la Dirección General de Farmacia y Productos Sanitarios, de manera que todo ensayo clínico con medicamentos que se realiza en España, además de seguir el procedimiento administrativo establecido por la ley<sup>27</sup>, es registrado en dicha base de datos. No obstante, para que esta medida contribuya a soslayar el potencial impacto del sesgo de publicación, la base de datos debería mantenerse actualizada, completada con información relevante (lo que implica, a su vez, un seguimiento de los ensayos clínicos autorizados) de forma suficientemente exhaustiva y explotada adecuadamente.

#### *Informes estructurados de los ensayos clínicos (recomendaciones del grupo de trabajo Consolidated Standards of Reporting Trial [CONSORT])*

Es importante distinguir entre la calidad de un ensayo clínico y la de su informe o publicación. Es posible que un ensayo clínico deficiente con numerosos sesgos sea publicado de forma impecable. Por contra, un ensayo clínico correctamente diseñado y realizado puede ser publicado de forma muy deficiente. No obstante, los estudios desarrollados sugieren que la calidad de los informes de los ensayos clínicos y la de su diseño y ejecución están muy correlacionadas, por lo que es posible establecer unas guías que faciliten no sólo la correcta publicación de los ensayos clínicos por sus autores, sino también su valoración por parte de los lectores u observadores externos.

Son varias las guías o directrices publicadas hasta la fecha a tal efecto, algunas de ellas en forma de prácticos listados (*checklists*), pero su impacto en la mejora de los informes de los ensayos clínicos ha sido limitada. De entre todas las diversas iniciativas<sup>28-46</sup>, cabe destacar la conocida como CONSORT (Consolidated Standards of Reporting Trials)<sup>47-52</sup>, inicialmente adoptada por la revista *JAMA* en 1996 y posteriormente por muchas otras revistas de prestigio (*British Medical Journal*, *The Lancet*, *Canadian Medical Association Journal*, entre otras). Ésta establece una lista de 21 ítems relacionados tanto con los aspectos metodológicos como de los resultados y su discusión e identifica los puntos claves de la información necesarios para evaluar tanto la validez interna como externa de un informe (anexo I). Un aspecto relevante en relación a los ítems propuestos es que se basan en datos empíricos acerca de su potencial impacto sobre la estimación del efecto de las intervenciones, de manera que su omisión podría resultar en una estimación sesgada de aquél. Así mismo, se propone un diagrama de flujo para proporcionar información detallada acerca del progreso de los pacientes a lo largo del estudio, en los distintos grupos de intervención que configuran el ensayo clínico (fig. 1). Finalmente, se propone un formato estándar para la publicación de los ensayos clínicos que incluye la combinación de cinco nuevos apartados en el texto del informe (protocolo, asignación del tratamiento, enmascaramiento, curso y seguimiento de los pacientes y análisis), así como el uso de listados durante el proceso de revisión editorial, como requisito previo necesario para su publicación.

ANEXO I  
PROPUESTA CONSORT: LISTADO (*CHECK-LIST*) DE 21 ASPECTOS ESENCIALES  
PARA LA EVALUACIÓN DE LOS ENSAYOS CLÍNICOS

<i>Título</i>	<i>Apartado</i>	<i>Descripción</i>
Título		Identificar el estudio como un ensayo aleatorizado
Resumen		Utilizar un formato estructurado
Introducción		Especificar las hipótesis, los objetivos clínicos así como los análisis de subgrupos o de covarianza definidos previamente
Métodos	Protocolo	<p>Describir:</p> <p>La población de estudio así como los criterios de inclusión y exclusión</p> <p>Las intervenciones previstas y su calendario de ejecución</p> <p>Las variables de resultado principal y secundaria/s, el valor de la mínima diferencia que se considerará clínicamente relevante e indicar cómo se calculó el tamaño de la muestra</p> <p>Justificación y explicitación del análisis estadístico propuesto, detallando los principales análisis comparativos y si éstos se efectuaron sobre las bases de intención de tratar</p> <p>Criterios establecidos <i>a priori</i> para la interrupción del estudio</p>
	Asignación	<p>Describir:</p> <p>Unidad de aleatorización (p. ej., sujeto, grupo, geográfico, etc.)</p> <p>Procedimiento utilizado para la asignación aleatoria</p> <p>Métodos para encubrir la asignación y fases de su ejecución</p> <p>Método para garantizar la separación entre el generador y el ejecutor de las asignaciones</p>
	Enmascaramiento	Describir el mecanismo (p. ej., cápsulas, tabletas); similitud de las características del tratamiento (p. ej., aspecto, sabor); control de las asignaciones resultantes (ubicación de los códigos durante el ensayo y condiciones para su violación), y evidencia de que el enmascaramiento entre los participantes, los investigadores, los evaluadores y los responsables del análisis de los datos fue efectivo
Resultados	Flujo y seguimiento de los participantes	Incluir un diagrama que resuma el flujo de los participantes a lo largo del estudio, su número y el cronograma de la asignación aleatoria y de las intervenciones y medidas efectuadas en cada grupo de estudio
	Análisis	<p>Especificar la medida estimada del efecto de la intervención para cada variable de resultado principal y secundaria, incluyendo su estimación puntual y una medida de precisión (intervalo de confianza)</p> <p>Presentar los resultados en números absolutos, siempre que sea posible (p. ej., 10/20 en lugar de 50%)</p> <p>Presentar el resumen de los datos así como los estadísticos descriptivos e inferenciales con el suficiente detalle para permitir análisis alternativos y/o su reproducción</p> <p>Describir la distribución de las variables de pronóstico en cada uno de los grupos de estudio y especificar si se ha efectuado algún tipo de ajuste</p> <p>Describir cualquier desviación del protocolo de estudio y su justificación</p>
Comentarios		<p>Interpretación de los hallazgos del estudio, teniendo en cuenta las posibles fuentes de sesgo y de imprecisión (validez interna), y discusión de la validez externa, incluyendo medidas cuantitativas apropiadas cuando sea posible</p> <p>Interpretación general de los datos a la luz de la evidencia científica disponible</p>

Traducido y adaptado de *JAMA* 1996; 276: 637-639.



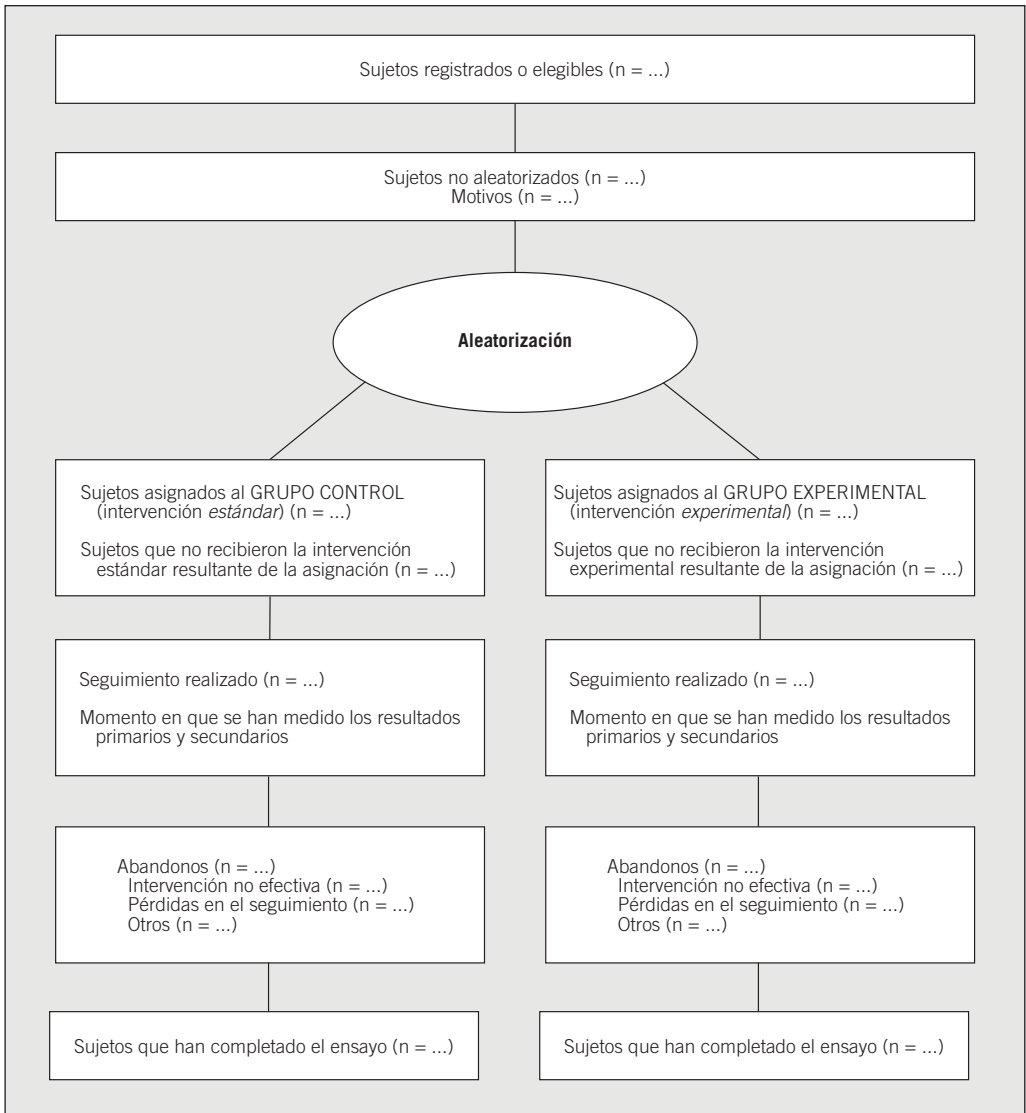


Fig. 1. Propuesta CONSORT: diagrama de flujo de los sujetos a lo largo del ensayo clínico.

Esta propuesta podría ser adoptada fácilmente por parte de las revistas biomédicas españolas, así como recomendada por las autoridades sanitarias y los comités éticos de investigación clínica, lo que redundaría en una mejora sustancial de la calidad tanto en el diseño y realización como en la comunicación y evaluación de los resultados de los ensayos clínicos llevados a cabo en España.

*Papel de los comités éticos de investigación clínica: salvaguarda de la libertad de los investigadores e instituciones e incentivación para la publicación de los resultados de los ensayos clínicos por parte de los investigadores*

La no publicación de los resultados tiene una más amplia implicación que las regulaciones le-

gales debieran tener en cuenta. Son varias las razones esgrimidas por muchos autores en el sentido de que la publicación de los resultados de investigación constituye un imperativo ético ineludible<sup>53</sup> cuya omisión constituye una mala práctica científica por omisión de un deber, ya que los sujetos deciden dar su consentimiento para participar en un ensayo clínico (aceptando voluntariamente sus riesgos potenciales y molestias inherentes) en la comprensión de que de éste se derivará algún tipo de beneficio social. En la tabla I se presenta un extracto del redactado de una hoja de información al paciente que bien podría ser considerada como representativa de la mayoría de estos documentos que son presentados a los comités éticos españoles. Por esta razón, si queremos ser honestos con estos pacientes, deberíamos poder asegurarles que los datos serán analizados sin sesgos y que los hallazgos serán del dominio público y, por tanto, estarán disponibles para posibles interesados, lectores o revisores. Tanto los propios investigadores como los comités éticos tienen el deber y el poder para hacer posible este compromiso<sup>54</sup>. Como algunos autores han afirmado, la garantía de publicación de los resultados de los ensayos clínicos debería ser considerada como un paso importante e ineludible para prevenir la violación de la dignidad humana de los sujetos participantes en investigación biomédica<sup>55</sup>.

Así mismo, la publicación de los resultados de los ensayos clínicos constituye tanto una prueba de la conducta ética del investigador como una medida indirecta del rigor de los comités éticos que los han aprobado y de las instituciones que éstos representan. Los comités éticos deberían insistir, pues, en la necesidad de que exista una intención manifiesta y explícita de publicar los resultados de todos los estudios experimentales con seres humanos, sea cual sea la dirección que tomen sus resultados<sup>56</sup>.

#### *Acuerdos con la industria (contratos de investigación entre el centro y el promotor)*

Un modo evidente e inmediato en el que pueden hacer efectivo este objetivo tanto los comités éticos de investigación clínica locales como los centros respectivos a los que éstos representan es mediante los acuerdos que, tras la aprobación del protocolo de ensayo clínico, se establecen entre la dirección del centro y el promotor, referenciados por el oportuno contrato. Puesto que no existe un único modelo de contrato aceptable para todos los centros, y puesto que la ley no incide directamente en sus contenidos, debería te-

TABLA I  
EXTRACTO DE UNA HOJA DE  
INFORMACIÓN AL PACIENTE:  
UN EJEMPLO ESTÁNDAR

#### *Riesgos y beneficios del estudio*

Al tomar parte en el estudio usted también beneficiará a otros pacientes que en el futuro desarrollen "la enfermedad de interés", ya que podrán beneficiarse de los datos generados en este estudio. Por tanto, al ayudar a establecer si el "fármaco en estudio" es activo contra la "enfermedad de interés", además de ayudarse a sí mismo, ayudará a otras personas con "la enfermedad de interés"

nerse sumo cuidado a la hora de redactar y aprobar determinadas cláusulas, específicamente las relacionadas con los compromisos que adquieren las partes implicadas en la realización del ensayo clínico, la propiedad de los datos y su confidencialidad y las condiciones para la publicación, entre otras (tabla II).

De esta forma, los contratos deberían salvaguardar la libertad de los investigadores, así como de las instituciones a las que éstos representan para la publicación de los resultados respetando, obviamente, los derechos que asisten también al promotor. Entre éstos se encuentran la necesidad de obtener el consentimiento previo y por escrito del promotor antes de la divulgación y/o publicación de los resultados, la remisión de un borrador al promotor para sus comentarios oportunos antes de la entrega de la versión final para su edición, la publicación de los resultados globales (en los estudios multicéntricos) antes que la referida a alguno de los centros participantes, entre otros. Además, dichas cláusulas deberían garantizar que las diversas opiniones, en caso de existir discrepancia entre las partes, quedaran debidamente reflejadas en el documento. Estas cláusulas deberían ser aplicadas también a la información obtenida en los ensayos clínicos no finalizados o suspendidos antes de su finalización.

Por otra parte, resulta claro que los contratos deberían tener en cuenta debidamente las particularidades de determinados ensayos clínicos (en especial aquéllos con nuevas moléculas) en los que deben extremarse las garantías de confidencialidad en el manejo de la información proporcionada y generada para preservar el legítimo interés del promotor. La industria, a menudo, ha llevado a cabo inversiones millonarias en investigación que podrían verse seriamente perjudica-

TABLA II  
 DIVERSOS REDACTADOS REFERIDOS AL APARTADO DE PUBLICACIONES,  
 EXTRAÍDOS DE CONTRATOS REALES DE ENSAYOS CLÍNICOS

<p>“ El investigador es libre de utilizar los resultados derivados de este estudio con fines científicos, aunque deberá comentar con el (<i>Promotor</i>) cualquier publicación antes de su edición. El (<i>Promotor</i>) reconoce el derecho de los investigadores a publicar los resultados del estudio tras la finalización del mismo. Los investigadores se comprometen a remitir al (<i>Promotor</i>) un borrador del manuscrito para publicar o un resumen para su aprobación antes de la entrega de la versión final para su edición. Estos documentos se revisarán con celeridad y no se negará la aprobación de forma injustificada”</p> <p>“ En caso de que existan diferencias entre las opiniones del investigador y las de promotor, en la discusión de los datos se expondrá la interpretación de ambas partes. En caso de que el investigador no desee publicar los resultados del estudio, el promotor es libre de utilizar los datos del estudio para su publicación. En este caso, el promotor remitirá el manuscrito al investigador quien tendrá derecho a insistir en que su interpretación de los resultados quede adecuadamente reflejada”</p> <p>“ Los resultados del ensayo clínico serán de la exclusiva propiedad del Promotor. Aun</p>	<p>así, el investigador principal podrá presentarlos en una reunión científica apropiada y/o publicarlos en una revista de reconocido prestigio, comprometiéndose a suministrar al Promotor una copia del manuscrito u original, con la suficiente antelación, a efectos de que éste tenga oportunidad de conocer dicha información o material informativo para la realización de sus comentarios sobre el contenido de tales comunicaciones”</p> <p>“ En circunstancias excepcionales, si el Promotor así lo solicita, con el fin de asegurar apropiadamente la protección de invenciones o desarrollos derivados del ensayo, el investigador principal acepta retrasar la presentación de la publicación propuesta, durante un plazo no superior a 6 meses”</p> <p>“ El (<i>laboratorio promotor</i>) está en contra de la publicación de los datos de un centro individual en caso de ensayos multicéntricos ya que el esfuerzo fue diseñado para analizarlo utilizando los datos de todos los centros. De todas maneras, no se podrá realizar ninguna publicación hasta que todos los centros hayan completado el ensayo”</p>
--	---

das por otras cuestiones ajenas a la obtención de unos resultados eventualmente negativos en la investigación.

#### *Modificaciones legislativas favorables a la publicación de los resultados de los ensayos clínicos*

Las legislaciones actuales no contemplan la obligación por parte del investigador o promotor de un ensayo clínico para que éste publique los resultados, incluso cuando éstos no apoyan sus propios intereses comerciales, profesionales o económicos. Esto concierne de forma particular, aunque no exclusiva, a los ensayos clínicos promovidos por la industria farmacéutica, que constituyen más del 80 % en España. Sin desmerecer en absoluto el importante esfuerzo que la industria ha venido realizando en este campo y su valiosa contribución a la generación de conocimiento relevante, a nadie escapa los posibles

intereses que pueden conducir a una deliberada ocultación de la información generada.

En España, la legislación<sup>27</sup> establece los requisitos y condiciones que deben cumplirse para que el diseño, ejecución y comunicación de los resultados de los ensayos clínicos –de producirse– permitan asegurar que los datos son fiables y que se ha preservado el respeto a los postulados éticos. No obstante, dichas referencias son genéricas apelando a las normas de buena práctica clínica, sin establecerse una obligación para la comunicación de los resultados. Sí queda establecida claramente la obligatoriedad de la elaboración de un informe final que debe ser remitido a la autoridad sanitaria y a los Comités Éticos de Investigación Clínica (CEIC) a la finalización del mismo, así como informes anuales sobre la marcha del estudio en caso de que su duración sea superior a un año (artículo 32).

Una posible medida normativa al respecto sería la catalogación de información no confidencial

de los informes de seguimiento y, en especial, del informe final de un ensayo clínico. Tratándose así de un documento público de cumplimiento obligatorio, se garantizaría el acceso libre y la difusión de los resultados generados por los ensayos clínicos realizados en España, siempre y cuando, claro está, los informes se realicen, pero ésta sí es una obligación inexcusable del investigador (y del CEIC de exigirlo) contemplada en la legislación actual.

Otro aspecto relevante establecido por la legislación y que debiera ser objeto de especial atención es la necesidad de justificar convenientemente las propuestas de investigación que son sometidas para valoración, lo que implica un escrutinio adecuado del conocimiento existente al respecto para minimizar los riesgos de los sujetos participantes (tabla III). Podríamos entender como riesgos inaceptables los que son injustificables en función del beneficio esperado del estudio o también los innecesarios derivados de la investigación redundante. Este requerimiento sólo podrá ser garantizado en la medida que se conozcan los resultados de los ensayos clínicos.

Es de prever que en un futuro inmediato se producirán importantes cambios legislativos que afectarán a los países miembros de la Unión Europea. Dichos cambios vendrán determinados por la aprobación de una futura Directiva del Parlamento Europeo que pretende armonizar las legislaciones locales referentes a la realización de ensayos clínicos con medicamentos. Esta directiva, o sus sucesivas modificaciones, supone una oportunidad única para introducir a nivel europeo cambios en el sentido hasta aquí señalado.

#### *Modificaciones en los documentos o guías de referencia en la investigación*

Ni las normas de buena práctica clínica ni la Declaración de Helsinki u otros documentos afines (como documentos de referencia sobre la conducta éticamente aceptable en la investigación biomédica), a pesar de fundamentarse en los principios básicos de la bioética, hacen referencia alguna a la obligación ética de publicar los resultados de la investigación. No obstante, hoy día resultaría difícilmente justificable la no modificación de dichos documentos de referencia abogando por la necesidad ética de la libre comunicación de los resultados de la investigación en seres humanos.

Algunos autores han propuesto introducir modificaciones en las normas de buena práctica clínica que afectan también a la monitorización de los ensayos clínicos, como la creación de comi-

### TABLA III REFERENCIAS A LA JUSTIFICACIÓN DE LOS ENSAYOS CLÍNICOS EN EL REAL DECRETO 561/1993

“ Sólo se podrán realizar ensayos clínicos cuando se cumplan todos los principios siguientes:... b) el estudio se base en los conocimientos actuales disponibles, la información buscada suponga, presumiblemente, un avance en el conocimiento científico del ser humano o para mejorar su estado de salud y su diseño minimice los riesgos para los sujetos participantes en el mismo” (artículo 10.3)

“ El CEIC... evaluará la idoneidad del protocolo en relación con los objetivos del estudio, su eficiencia científica (la posibilidad de alcanzar conclusiones válidas, con la menor exposición posible de sujetos) y la justificación de los riesgos y molestias previsibles, ponderadas en función de los beneficios esperados para los sujetos y la sociedad” (artículo 42.1)

CEIC: Comité Ético de Investigación Clínica.

tés de monitorización de los datos independientes, así como un análisis de los datos por parte de instituciones independientes del promotor del ensayo clínico<sup>57</sup>. Se destaca también la obligación moral de llevar a cabo un análisis intermedio de los datos. Por otra parte, se propone una mayor accesibilidad por parte de los investigadores participantes, así como por instituciones científicas independientes a los datos de los ensayos clínicos.

#### *“Amnistía” de ensayos clínicos no publicados*

Dadas las importantes consecuencias de los ensayos clínicos no publicados, los editores de aproximadamente 100 revistas médicas internacionales se han puesto de acuerdo para anunciar lo que se ha denominado “amnistía de ensayos clínicos no publicados”<sup>58</sup>. Mediante esta iniciativa se daba una oportunidad a los investigadores para hacer públicos los resultados de ensayos clínicos previamente no publicados o bien sólo comunicados en forma de resumen, lo que permite que contribuyan a la fundamentación científica de la atención sanitaria. Así, a través de todas estas revistas se proporciona una ficha (*unreported-trial registration form*) que puede ser cumplimentada y enviada a la oficina editorial por cual-

quier persona capaz de proporcionar la información solicitada, aunque no dispusiera de los datos del estudio. La información enviada es hecha pública mediante una lista de ensayos clínicos a través de Internet. De este modo, los potenciales revisores interesados en obtener una información detallada de estos estudios pueden dirigirse directamente a los investigadores para solicitarla.

*Desarrollo de estrategias nacionales e internacionales por parte de los organismos públicos responsables de la política científica*

Se está concretando una importante iniciativa por parte de dos Estados europeos y que bien podría generalizarse pronto a muchos otros países, cuando menos de la Unión Europea, en el marco del proceso actual de armonización y convergencia de las legislaciones locales. Así, tanto el Danish Research Ethics Committee System como el United Kingdom Medical Research Council, conscientes de la importancia de la revisión exhaustiva de la bibliografía científica y la influencia que deben tener los conocimientos científicos ya existentes en el proceso de evaluación ética de los estudios experimentales con seres humanos, han hecho públicas una serie de recomendaciones al respecto entre las que destaca la necesidad de llevar a cabo, por parte del equipo investigador, una revisión exhaustiva de toda la bibliografía relevante sobre la cuestión en estudio a través de las principales bases de datos bibliográficas disponibles como requisito previo a su valoración por un comité ético. Esto es especialmente relevante cuando el comité debe juzgar una propuesta de investigación original o bien cuando debe decidir sobre la oportunidad de utilizar placebo en lugar de otras alternativas terapéuticas conocidas en el grupo control. Así mismo, en caso de que aparezcan nuevos resultados en la bibliografía nacional o internacional mientras se está llevando a cabo el estudio y que alteran las bases éticas o la justificación de la propuesta de investigación original (p. ej., un estudio que contesta suficientemente la pregunta que el estudio en marcha trata de contestar, o que al demostrar claramente la eficacia de un nuevo tratamiento pone en cuestión la legitimidad de la utilización de un grupo placebo), dicho proyecto debe ser nuevamente evaluado por el comité ético a la luz de la nueva evidencia. Otro aspecto de especial importancia es el derecho por parte de los sujetos participantes en la investigación a ser informados de los cambios que se vayan produciendo *a posteriori*. Así, si los resultados apuntan, por ejemplo, hacia la necesidad

de un cambio en la dosificación o en la duración del tratamiento en la cohorte de intervención, los investigadores deben evaluar juntamente con el comité ético la necesidad de informar de dichos cambios a los pacientes que hayan terminado el estudio, así como proponer el cambio de régimen terapéutico correspondiente a los pacientes todavía en estudio. Finalmente, se establece la obligación moral por parte del equipo investigador de informar a los sujetos participantes en la investigación de todas las cuestiones relevantes que pudieran surgir a lo largo del desarrollo del estudio. Esta información incluye, para los ensayos clínicos finalizados, la comunicación por carta a todos los participantes de los resultados finales del estudio, del tratamiento asignado a cada paciente en concreto y de las posibles consecuencias prácticas de los resultados para los mismos sujetos. En el caso de los ensayos clínicos interrumpidos prematuramente antes de la obtención de resultados, implica la comunicación a todos los sujetos previamente participantes de dicha circunstancia.

Así mismo, la propuesta británica destaca la obligación por parte del investigador de asegurar que los resultados del estudio son analizados, comunicados y diseminados a través de revistas de prestigio (aquellas con un sistema de *peer-review*), independientemente de cuáles sean sus resultados.

## Conclusiones

Los ensayos clínicos constituyen una fuente esencial de conocimiento científico relevante que contribuye, de forma muy significativa, al progreso de la medicina y de los sistemas sanitarios, especialmente en cuanto a la oferta de tratamientos disponibles. Así, los ensayos clínicos permiten establecer la eficacia y la seguridad de nuevos fármacos hasta el momento no disponibles, o bien explorar nuevas indicaciones o condiciones de uso de los ya existentes, todo lo cual redundará en un incremento del arsenal terapéutico disponible, así como en su mejor utilización. El único requisito imprescindible para ello es su oportuna difusión mediante la publicación de los resultados alcanzados, lo que no siempre ocurre. Son varias las razones identificadas por las que numerosos ensayos clínicos no son publicados, especialmente los mal denominados “ensayos clínicos” (la utilización habitual de dicho adjetivo parece inducir erróneamente a enjuiciar la relevancia de los resultados y no sólo la dirección de éstos), dando lugar al fenómeno conocido como sesgo de publicación.

Por otra parte, son varios los actores que intervienen en su realización (promotor, investigadores, pacientes, comités éticos, centros sanitarios y autoridad sanitaria), todos necesarios y con intereses legítimos, aunque no siempre necesariamente coincidentes. No obstante, dado el vacío legal existente, así como la falta de mecanismos y propuestas al respecto, la prerrogativa para la publicación de sus resultados recae de forma exclusiva en la figura del promotor (“quien paga manda”). Sin embargo, se detecta un interés creciente en la salvaguarda de los intereses de las demás partes, lo que está llevando a la formulación de propuestas que, desde los diferentes frentes y a distintos niveles, contribuyan a soslayar dicho déficit. Algunas de ellas ya se están materializando mediante la adopción de disposiciones legales o de normativas de consenso, mientras que otras todavía se encuentran en una fase incipiente de discusión y sólo pretenden contribuir a generar un debate sobre algunos aspectos no cuestionados hasta fechas recientes.

En resumen, mucho se mueve en el panorama de la investigación clínica en todo el mundo y en el de las publicaciones de mayor prestigio. Sin duda alguna, estos esfuerzos pretenden devolver todo el sentido social que tiene, o que debería tener, la investigación experimental en humanos. En España no andamos sobrados de criterios y sobre todo, de acciones, que nos permitan permanecer al margen de estas iniciativas. ¿Por qué no nos sumamos a las mismas, mejorándolas si es posible?

## BIBLIOGRAFÍA

- Office of Technology Assessment. The impact of randomized clinical trials on health policy and medical practice. Background paper. Washington D.C.: US Congress, 1983.
- Porta M, Ibáñez I, Carné X, Laporte JR. Principios del ensayo clínico. *Med Clin (Barc)* 1983; 80: 768-770.
- Hulley SB, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma, 1993.
- Bakke OM, Carné X, García Alonso F. Ensayos clínicos con medicamentos. Fundamentos básicos, metodología y práctica. Barcelona: Doyma, 1994.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Medicina basada en la evidencia. Cómo ejercer y enseñar la MBE. Madrid: Churchill Livingstone, 1997.
- Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1.885-1.889.
- Ioannidis JP, Cappelleri JC, Sacks HS, Lau J. The relationship between study design, results, and reporting of randomized clinical trials HIV infection. *Control Clin Trials* 1997; 18: 431-444.
- Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *Br Med J* 1997; 315: 640-645.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-872.
- Moscato R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Acad Emerg Med* 1994; 1: 267-271.
- Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993; Apr 28; Doc N.º 50.
- Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. *Control Clin Trials* 1987; 8: 33-53.
- Chalmers I, Adams M, Dickersin K, Hetherington J, Tarnow-Mordi W, Meiner C et al. A cohort study of summary reports of controlled trials. *JAMA* 1990; 263: 1.401-1.405.
- Chalmers TC, Frack CS, Reitman D. Minimizing the three stages of publications bias. *JAMA* 1990; 263: 1.392-1.395.
- Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J Roy Stat Soc A* 1988; 151: 419-463.
- Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989; 81: 107-115.
- Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989; 84: 381-392.
- Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987; 6: 11-29.
- Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986; 4: 1.529-1.541.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-872.
- Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. *JAMA* 1992; 267: 374-378.
- Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993 Apr 28; Doc N.º 50.
- National Institute of Health inventory of clinical trials: fiscal year 1979.
- Meinert CL. Toward prospective registration of clinical trials. *Control Clin Trials* 1988; 9: 1-5.
- Dickersin K. Report from the panel on the case for registers of clinical trials at the eighth Annual Meeting of the Society for Clinical Trials. *Control Clin Trials* 1988; 9: 76-81.
- Dickersin K. Why register clinical trials? *Control Clin Trials* 1992; 13: 170-177.
- Real Decreto 561/1993, de 16 de abril, por el que se establecen los requisitos para la realización de ensayos clínicos con medicamentos.

28. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Technol Assess Health Care* 1996; 12: 195-208.
29. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4: 942-951.
30. Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. *Cancer Treatment Reports* 1985; 69: 1-3.
31. Grant A. Reporting controlled trials. *Br J Obstet Gynaecol* 1989; 96: 397-400.
32. Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol* 1983; 1: 164-169.
33. Mosteller F, Gilbert JP, McPeck B. Reporting standards and research strategies for controlled trials. *Control Clin Trials* 1980; 1: 37-58.
34. Gardner MJ, Machin D, Campdell MJ. Use of checklists in assessing the statistical content of medical studies. *Br Med J* 1986; 292: 810-812.
35. Gore SM, Jones G, Thompson SG. The Lancet's statistical review process: areas for improvement by authors. *Lancet* 1992; 340: 100-102.
36. 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: 125-128.
37. Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Assoc J* 1964; 90: 565-569.
38. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987; 317: 426-432.
39. Altman DG, Doré CJ. Randomization and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-153.
40. Gotsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10: 31-56.
41. Solomon MJ, McLeod RS. Clinical studies in surgical journals: have we improved? *Dis Colon Rectum* 1993; 36: 43-48.
42. 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: 408-412.
43. Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; 106: 598-604.
44. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990; 113: 69-76.
45. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*. *Can Med Assoc J* 1994; 150: 1.611-1.615.
46. Rennie D. Reporting randomized controlled trials: an experiment and a call for responses from readers. *JAMA* 1995; 273: 1.054-1.055.
47. Begg C, Cho M, Eastwood S et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996; 276: 637-639.
48. Fletcher RH, Fletcher SW. Clinical research in general medical journals: a 30-year perspective. *N Engl J Med* 1979; 301: 180-183.
49. Rennie D. How to report randomized controlled trials: the CONSORT statement. *JAMA* 1996; 276: 649.
50. Altman DJ. Better reporting of randomised controlled trials: the CONSORT statement. *Br Med J* 1996; 313: 570-571.
51. McNamee D, Horton R. Lies, damn lies, and reports of randomised controlled trials. *Lancet* 1996; 348: 562.
52. Huston P, Hoey J. CMAJ endorses the CONSORT statement. *Can Med Assoc J* 1996; 155: 1.277-1.279.
53. Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990; 263: 1.405-1.408.
54. Munro AJ. Publishing the findings of clinical research. *Br Med J* 1993; 307: 1.340-1.341.
55. Pearn J. Publication: an ethical imperative. *Br Med J* 1995; 310: 1.313-1.315.
56. Whitaker VB, Whitaker R Jr. Ethics: issues of concern. *Insight* 1991; 16: 30-31.
57. Jongen PJ. Data handling in clinical trials: an ongoing debate. *Mult Scler* 1995; 1 (Supl 1): 60-63.
58. Horton R. Medical editors trial amnesty. *Lancet* 1997; 350: 756.

---

# Evaluación de la calidad de los ensayos clínicos: justificación, utilidad e inconvenientes

---

Fernando J. García López<sup>a</sup>, Sonia Gutiérrez Bezón<sup>b</sup>,  
Inés Galende Domínguez<sup>c</sup> y Cristina Avendaño Solá<sup>d</sup>

<sup>a</sup>Unidad de Epidemiología Clínica y <sup>d</sup>Servicio de Farmacología Clínica.  
Hospital Universitario Clínica Puerta de Hierro. Madrid. <sup>b</sup>Unidad de Investigación Clínica.  
Sanofi Winthrop. Madrid. <sup>c</sup>Unidad de Ensayos Clínicos. Consejería de Salud. Comunidad de Madrid.

Se considera que el ensayo clínico controlado con distribución aleatoria es el tipo de estudio epidemiológico que presenta menores errores sistemáticos o sesgos y constituye el fundamento más fiable para apoyar la eficacia de las intervenciones terapéuticas o preventivas, ya directamente por la información proporcionada por un solo ensayo<sup>1,2</sup> o indirectamente a través de las revisiones sistemáticas de distintos ensayos<sup>3</sup>. Sin embargo, las distintas fases de un ensayo clínico, su diseño, su ejecución, el análisis de los resultados y su difusión, están sometidas a múltiples sesgos potenciales que pueden comprometer su validez. Esto ha dado lugar a una amplia bibliografía sobre la calidad de los ensayos clínicos desde enfoques diversos.

## Calidad como validez interna

---

Antes de hablar de la calidad de los ensayos habrá que especificar de qué calidad se está hablando. Queremos referirnos a la validez científica del propio ensayo, al grado en que su diseño y ejecución impiden los errores sistemáticos y se acercan a la "verdad" de la cuestión<sup>4</sup>, dicho en otras palabras, la validez entendida como la eliminación de sesgos en los procesos de diseño, ejecución y análisis, el grado en el que las conclusiones de un estudio reflejan correctamente lo que sucedió en realidad. Nos referimos a la definición de validez interna de Last: "los grupos en estudio se seleccionan y comparan de tal modo que las diferencias que se observan entre ellos con respecto a las variables dependientes del estudio, dejando aparte el error aleatorio, pueden achacarse exclusivamente a la hipótesis planteada en la investigación"<sup>5</sup>. También nos referiremos a la calidad de la descripción del ensayo clínico en la publicación, es decir, en qué medida la descripción refleja adecuadamente todas y cada una

de las fases del ensayo. En realidad, no siempre es fácil separar la validez interna del ensayo de la calidad de su publicación, sobre todo cuando en muchas ocasiones la evaluación de la validez científica debe hacerse con la información contenida en la publicación del ensayo. No vamos a entrar en otras posibles interpretaciones de la calidad, como pueden ser la validez externa o aplicabilidad de los resultados de los ensayos clínicos a otros pacientes distintos de los que participaron, su precisión o la medida de la probabilidad de obtener errores aleatorios que se refleja en la amplitud del intervalo de confianza alrededor del cálculo del efecto, la relevancia clínica de la cuestión abordada en el ensayo o, por último, la adecuación del ensayo a los patrones éticos exigibles en la investigación científica médica.

## Posibles errores sistemáticos en los ensayos clínicos aleatorizados

---

Existen numerosas pruebas empíricas de sesgos relacionados con diversos aspectos del diseño, ejecución y análisis de los ensayos (tabla I). En cuanto a la asignación de los participantes a los distintos grupos, la distribución aleatoria reduce el sesgo de selección<sup>6,7</sup>; la distribución no aleatoria en sentido estricto o pseudoaleatoria, como el reparto alternante, por fechas de nacimiento, por día de la semana, por números de historia o de seguridad social puede dar lugar a grupos no comparables<sup>8,9</sup> y sus resultados podrían favorecer al grupo del tratamiento experimental con respecto al tratamiento convencional<sup>10</sup>; la falta de ocultación del reparto aleatorio a los pacientes o a los médicos implicados en el ensayo, es decir, el hecho de que puedan conocer a qué grupo corresponde la siguiente asignación como, por ejemplo, cuando se utiliza una tabla de números aleatorios expuesta, favorece también al grupo del tratamiento experimental en los resultados finales del ensayo<sup>11-13</sup>.



TABLA I  
POSIBLES ERRORES SISTEMÁTICOS  
EN LOS ENSAYOS CLÍNICOS CON  
DISTRIBUCIÓN ALEATORIA DE LOS  
QUE HAY PRUEBAS EMPÍRICAS

---

Asignación a los grupos de tratamiento
Asignación no aleatoria
Asignación pseudoaleatoria
Falta de ocultación de la asignación
Enmascaramiento
Falta de enmascaramiento o
enmascaramiento incompleto del
observador
Falta de enmascaramiento o
enmascaramiento incompleto
del paciente
Análisis estadístico no enmascarado
Seguimiento de los participantes
Insuficiente descripción de los retirados
y de los abandonos
Análisis estadístico
Análisis que no respeta la asignación
a los tratamientos
Aplicación de técnicas estadísticas
incorrectas
Análisis de subgrupos sobrevalorados
Otros aspectos del diseño
Uso de variables “subrogadas” en lugar
de variables clínicas
Ensayos cruzados
Otros aspectos
Insuficiente aplicación de sistemas de
monitorización y de garantía de calidad
Ensayos unicéntricos en lugar de
multicéntricos
Falta de apoyo financiero explícito
Publicación en revistas de menor difusión
Ensayos financiados por la industria
farmacéutica publicados en
suplementos de revistas médicas

---

En cuanto al enmascaramiento del observador y del paciente con respecto a la intervención efectuada, los ensayos no enmascarados tienden a alterar también la evaluación de los resultados en sentido favorable al grupo del tratamiento experimental<sup>10,12</sup>, al igual que el análisis estadístico no enmascarado de los datos<sup>14</sup>. El enmascaramiento es necesario, aunque la evaluación de la variable principal no pueda verse afectada por el conocimiento sobre la variable de resultado (mortalidad u otras variables no influenciadas por la percepción del observador) debido a que pueden intervenir otros factores, como tratamientos con-

comitantes, actitudes de los sujetos o las decisiones de retiradas de los participantes<sup>15</sup>. Se ha observado que en estudios con diseño doble ciego imperfecto, los pacientes convencidos de que toman placebo, al que consideran menos eficaz que el tratamiento experimental, describen mayor cantidad de síntomas, lo que puede sesgar los resultados<sup>16</sup>.

En cuanto al seguimiento de los participantes, la insuficiente descripción del desenlace de los participantes en un ensayo por grupos, incluyendo los retirados, los abandonos y las violaciones del protocolo, puede dar lugar a resultados equivocados<sup>17,18</sup>.

En lo que se refiere al análisis estadístico, un análisis que no compare los grupos según la asignación a los tratamientos, o con expresión más desafortunada “por intención de tratar”<sup>19</sup>, sobre todo cuando se pretende demostrar la falta de equivalencia entre los tratamientos en estudio, también puede arrojar resultados erróneos<sup>17-21</sup>; la aplicación de técnicas estadísticas incorrectas puede convertir en significativas al nivel de probabilidad alta prefijado (p. ej., menor de 0,05) asociaciones que con las técnicas adecuadas tienen valores de probabilidad alfa superiores<sup>22</sup>; los análisis de subgrupos múltiples hechos *a posteriori* pueden dar lugar a resultados desorientadores<sup>23-25</sup>; la descripción de una parte seleccionada de los resultados puede ocasionar sesgos, sobre todo cuando se estudian múltiples variables de resultado y no se había especificado la variable principal de antemano<sup>14</sup>.

Con respecto a otros aspectos del diseño de los ensayos, también existen pruebas empíricas de que el uso de variables “subrogadas”, es decir, sustitutas o intermedias, en lugar de variables con relevancia clínica directa, puede llevar a interpretaciones equivocadas<sup>26-28</sup> y de que los ensayos cruzados pueden sobrestimar los efectos del tratamiento<sup>29</sup>. También existen ciertas pruebas empíricas acerca de la influencia parcial de varios factores externos sobre los elementos relacionados con la validez interna de los ensayos clínicos. El cumplimiento de las normas de buena práctica clínica<sup>30</sup> garantiza que el ensayo se realice estrictamente de acuerdo con el protocolo<sup>31</sup> y que se sigan todos los procedimientos encaminados a garantizar la credibilidad de los datos obtenidos. La monitorización del ensayo y la realización de auditorías, herramientas consustanciales a las normas de buena práctica clínica, han revelado conductas que van desde irregularidades en la solicitud del consentimiento informado o en la recogida de datos, hasta la falsificación de datos en casos extremos, pasando por

la alteración en el sistema de distribución aleatoria o la falta de seguimiento estricto del protocolo<sup>13,32</sup>. También existen evidencias empíricas de que los ensayos multicéntricos, en oposición a los unicéntricos, aquellos en los que figura un estadístico entre sus autores, los que se han sometido a un proceso formal de evaluación previo a su ejecución, los publicados en las revistas más prestigiosas y aquellos en los que se declara una institución financiadora expresa presentan mejores indicadores de calidad<sup>33</sup>. Por otra parte, los ensayos que se publican en suplementos de revistas médicas presentan peor calidad metodológica que los ensayos que se publican en los números ordinarios de las mismas revistas<sup>34</sup>.

### **Medidas para evaluar la validez interna del ensayo clínico**

Ante tanta vulnerabilidad potencial, es lógico que hayan proliferado múltiples patrones de evaluación para juzgar la calidad de los ensayos clínicos, que podemos clasificar en tres tipos: las evaluaciones de los componentes aislados, las escalas y las listas-guía<sup>35</sup>. Las evaluaciones de los componentes aislados abordan el aspecto o los aspectos concretos de los ensayos que se consideran de mayor relevancia para garantizar su validez interna, de entre los que acabamos de ver. Las escalas y las listas-guía intentan reunir en un patrón común varios apartados relativos a la calidad global del ensayo, en el que las listas-guía intentan hacer una estimación cualitativa de la calidad mediante la inclusión de una serie de componentes individuales, mientras que las escalas pretenden resumir en una cifra numérica el grado de calidad que reúne el ensayo a partir de una serie de requisitos clave.

### **Evaluación de la validez interna mediante listas-guía**

Recientemente, Moher et al efectuaron una búsqueda exhaustiva de todas las listas-guía y escalas descritas para evaluar la calidad de los ensayos clínicos<sup>4</sup>. Encontraron 9 listas-guía, cuatro de ellas orientadas a la evaluación de la calidad del ensayo<sup>36-39</sup>, tres a la evaluación de la descripción en la publicación<sup>40-42</sup> y dos a la evaluación de ambos tipos de calidad<sup>43,44</sup>. En nuestro país, se ha publicado una lista-guía no incluida en la búsqueda de Moher, con dos variantes, una para la evaluación de ensayos clínicos publicados y otra para la de protocolos de ensayos clínicos<sup>45</sup>. En todos los casos, estas listas-guía se basaron en criterios aceptados, recogidos en algunos de

los libros de texto sobre ensayos clínicos<sup>8,9</sup>. En la mayoría de los casos se incluían en las listas-guía apartados referentes a la asignación de los participantes al grupo terapéutico, al enmascaramiento, al seguimiento y al análisis estadístico.

### **Evaluación de la validez interna mediante escalas**

En la misma revisión, Moher et al<sup>4</sup> encontraron 25 escalas descritas para la evaluación de la calidad de los ensayos clínicos, ya fuera la de la validez interna del propio ensayo, la de su descripción o de la calidad de ambas, y tanto de ensayos juzgados individualmente<sup>46-57</sup> como de ensayos que se van a incluir en una revisión sistemática<sup>10,14,58-65</sup>. Moher et al tampoco incluyeron una escala desarrollada en nuestro medio<sup>33</sup>. Sólo en seis escalas hubo una definición previa de la validez de constructo de la calidad, es decir, del grado en el que la escala se corresponde con los conceptos teóricos o constructos que afectan a lo que se pretende medir. Únicamente en una escala<sup>63</sup> (tabla II), realizada para ensayos sobre el dolor pero generalizable a todos los ensayos, la selección de los apartados no se hizo según criterios definidos previamente por consenso sino mediante las técnicas psicométricas habituales de desarrollo de escalas<sup>66</sup>, es decir, métodos formales para garantizar su validez aparente y la comprobación de la capacidad de la escala para distinguir entre ensayos considerados buenos, malos o escogidos al azar según la opinión de expertos independientes. En menos de la mitad de los casos se describió la concordancia interobservador. A pesar de estas dificultades metodológicas, en la mayoría de las escalas (en el 80 % o más) se incluyeron apartados sobre la asignación de los participantes a los grupos en estudio del ensayo, sobre el enmascaramiento del observador y del participante, y sobre el análisis estadístico, y en menos de la mitad de las escalas se incluyeron apartados sobre el seguimiento de los participantes. Después de la revisión de Moher et al, otros autores han utilizado una escala basada en criterios definidos en el manual de la Colaboración Cochrane<sup>67,68</sup>, aunque no apoyada por ésta, y que presenta una escasa concordancia<sup>69</sup>.

Además del propio contenido de la escala, las condiciones de su aplicación también han sido objeto de investigación. Cuando se ha estudiado la aplicación de la escala de forma enmascarada con respecto a los autores, a sus instituciones y a los resultados del trabajo, las puntuaciones son distintas, unas veces mayores y otras menores, y más concordantes<sup>63,70</sup>, pero las diferencias en-

TABLA II  
 INSTRUMENTO PARA MEDIR LA VEROSIMILITUD DE SESGOS EN LOS INFORMES  
 DE INVESTIGACIONES SOBRE EL DOLOR<sup>63</sup>

<p>La cumplimentación de este instrumento no debe hacerse como cuando hay que revisar un artículo para publicación. Puntuar un informe no debería durar más de 10 minutos y no hay respuestas correctas ni falsas. Por favor, lea el artículo e intente contestar a las siguientes preguntas (vea las instrucciones adjuntas):</p> <ol style="list-style-type: none"> <li>1. ¿Se describió el estudio como aleatorizado (esto incluye el uso de palabras como aleatorizadamente [<i>randomly</i>], distribución al azar [<i>random</i>] y aleatorización [<i>randomization</i>])?</li> <li>2. ¿Se describió el estudio como doble ciego?</li> <li>3. ¿Hubo una descripción de los abandonos y de los retirados?</li> </ol> <p><b>Puntuación de las preguntas:</b></p> <p>Por cada respuesta "sí" dad una puntuación de 1 punto y por cada respuesta "no" dad una puntuación de 0 puntos. No existen puntuaciones intermedias.</p> <p>Dad 1 punto adicional si:      Para la pregunta 1, se describió el método para producir la secuencia de aleatorización y era un <i>método apropiado</i> (tabla de números aleatorios, generados por ordenador, etc.)</p> <p>y/o:      Si para la pregunta 2 se describió el método para hacer el estudio doble ciego y si era un <i>método apropiado</i> (placebo con características idénticas, placebo activo, placebo idéntico [<i>dummy</i>], etc.)</p> <p>Quitad 1 punto si:      Para la pregunta 1, se describió el método de producir la secuencia de la aleatorización y era un <i>método inapropiado</i> (la asignación de los pacientes fue alterna o según la fecha de nacimiento, el número de historia, etc.)</p>	<p>y/o:      Para la pregunta 2, se describió el estudio como doble ciego pero el método de enmascaramiento fue <i>inapropiado</i> (p. ej., la comparación de pastillas con inyección sin doble placebo idéntico [<i>dummy</i>])</p> <p><b>Guías para la valoración:</b></p> <ol style="list-style-type: none"> <li>1. <i>Aleatorización</i>        Se considerará apropiado un método para producir la secuencia de aleatorización si permitió a cada participante tener la misma oportunidad de recibir cada intervención y los investigadores no pudieron predecir qué tratamiento se iba a asignar después. No deberían considerarse apropiados los métodos de asignación que utilizan la fecha de nacimiento, la fecha de ingreso, los números de historia o los métodos alternantes</li> <li>2. <i>Doble enmascaramiento</i>        Se considerará un estudio como doble ciego si se usa el término doble ciego (<i>double blind</i>). Se considerará apropiado el método si se declara que ni la persona que efectúa la valoración ni la que participa en el estudio pudieron identificar la intervención recibida, o, en su defecto, si se menciona el uso de placebos activos, de <i>placebos idénticos</i> o <i>dummies</i></li> <li>3. <i>Retirados y abandonos</i>        Deben describirse los participantes que se incluyeron en el estudio pero que no completaron el periodo de observación o que no se incluyeron en el análisis. Debe declararse el número y las razones de las retiradas en cada grupo. Si no hubo retiradas, debería decirse explícitamente en el artículo. Si no hay ninguna mención a las retiradas, a esta pregunta no se le debe calificar con ningún punto</li> </ol>
--	--

contradas parecen de una magnitud pequeña. Otros aspectos controvertidos en cuanto a la aplicación de las escalas son el número de revisores que hagan la evaluación, los procedimientos para lograr el acuerdo en el caso de que surjan discrepancias y el grado de conocimiento que los revisores deban tener acerca del objeto de estudio del ensayo clínico<sup>35</sup>.

¿Qué influencia tiene la diversidad de las escalas en la evaluación de la calidad de ensayos clínicos concretos?, ¿hasta qué punto la evaluación realizada mediante distintas escalas produce resultados discrepantes? Con respecto a estas cuestiones, los estudios que las han investigado ofrecen resultados discordantes. Detsky et al encontraron resultados prácticamente equivalentes

cuando ordenaron 18 ensayos clínicos aleatorizados sobre nutrición parenteral según su calidad medida con tres escalas distintas<sup>48</sup>. Por contra, Moher et al compararon las puntuaciones obtenidas al evaluar 12 ensayos clínicos con seis escalas publicadas y encontraron amplias discrepancias en las puntuaciones de los ensayos, hasta el punto de que hubo ensayos que obtuvieron del 23 al 74 % del total posible de la puntuación, según la escala, y al ordenar los 12 ensayos según las puntuaciones obtenidas con cada escala hubo dos ensayos que obtuvieron la mejor y la peor puntuación en dos escalas distintas<sup>35</sup>. También hubo discrepancia en los resultados cuando se evaluó la calidad de los ensayos clínicos que comparaban contrastes radiológicos<sup>71,72</sup>.

### **Evaluación de la validez interna de los ensayos clínicos individuales en las revisiones sistemáticas**

Las revisiones sistemáticas constituyen un campo en el que la calidad de los ensayos clínicos es determinante. Cuando una revisión sistemática intenta resumir cuantitativamente en una medida sumaria el efecto de una intervención (metaanálisis), parte de la presunción de que los resultados de los ensayos clínicos individuales difieren entre sí sólo por azar. Pero si en algunos ensayos incluidos en la revisión existen sesgos que comprometen su validez y, por lo tanto, sus resultados no son ciertos, esto desequilibrará las medidas sumarias del metaanálisis en una magnitud correspondiente a su grado de sesgo y al peso dado al ensayo sesgado en el conjunto del metaanálisis. Por ejemplo, un análisis cuantitativo de los resultados de ensayos de calidad variable puede dar lugar a conclusiones "falsamente positivas", que achacan erróneamente un efecto a una intervención si los estudios de peor calidad sobrestiman el efecto beneficioso de una intervención terapéutica o, por contra, "falsamente negativas", que descartan erróneamente un efecto de una intervención si las subestiman<sup>48</sup>. Puede ocurrir incluso que los resultados de todos los ensayos incluidos en una revisión sistemática concuerden pero que todos ellos adolezcan de sesgos semejantes, con lo que las conclusiones de la revisión sistemática deben ser tomadas con bastante cautela<sup>68</sup>. Este peligro ha dado lugar a la propuesta de distintos métodos para incorporar la evaluación de la calidad de los ensayos clínicos individuales en las revisiones sistemáticas. Naturalmente, las amenazas a la validez dependen del tipo de enfermedad y de intervención estudiada, que condicionan el grado en que puede haber

sesgos. Por ejemplo, el potencial para sesgos es menor en estudios que consideren la mortalidad como variable de resultado principal, que es una variable más resistente a sesgos de observación y no requiere tanto de diseños de enmascaramiento como otros ensayos en los que las variables de resultado puedan verse influidas por la percepción del sujeto o del observador, como es el caso de la percepción del dolor u otros síntomas.

### **Métodos para incluir la evaluación de la validez interna de los ensayos en una revisión sistemática**

Los métodos para incorporar la calidad metodológica en la revisión sistemática se pueden agrupar a grandes rasgos en cuatro<sup>68,73-76</sup>.

#### *Primer método: síntesis de la mejor evidencia disponible*

El primer método, basado en el concepto de la "síntesis de la mejor evidencia disponible"<sup>77,78</sup>, consiste en incluir en la revisión sólo aquellos ensayos que superen un cierto umbral mínimo de calidad. Cómo se define dicho umbral mínimo es, sin embargo, una cuestión en la que no hay acuerdo. En la mayor parte de las revisiones sistemáticas, sólo se incluyen aquellos ensayos con distribución aleatoria. Otros autores han propuesto, además, la exclusión de los ensayos no publicados como artículos completos, pero lo hace desaconsejable el riesgo que puede tener sobre la validez del metaanálisis el sesgo de publicación; es decir, la influencia que la índole de los resultados del estudio, si son favorables a la intervención experimental o no lo son, tiene sobre sus posibilidades de publicarse, aunque no existe unanimidad en esta cuestión<sup>79-85</sup>. La aplicación de un umbral de calidad basado en alguna escala, de modo que sólo se incluyan los ensayos que sobrepasan una determinada puntuación de la escala, plantea dos inconvenientes: en primer lugar, como hemos visto, los resultados pueden depender ampliamente de la escala escogida; en segundo lugar, el carácter arbitrario del umbral que se quiera fijar también puede modificar consiguientemente los resultados, ya que no hay pruebas empíricas en favor de ningún criterio mínimo global que debe cumplir un ensayo clínico para poder incluirlo en una revisión.

#### *Segundo método: relación entre la calidad de los ensayos y su variabilidad*

Un segundo método para evaluar la calidad metodológica consiste en explorar la relación

existente entre la calidad de los estudios y la variabilidad (heterogeneidad) encontrada en la revisión sistemática entre los resultados de los ensayos individuales. Una primera aplicación de este método es un análisis exploratorio de la relación entre los distintos valores del efecto en los estudios individuales y sus puntuaciones en la escala de calidad. En este sentido, se han propuesto técnicas gráficas, entre las que se encuentran los “gráficos en embudo”, que dibujan los valores del efecto de cada ensayo con respecto a sus tamaños de muestra, de modo que si presentan un aspecto asimétrico sugieren sesgos en los ensayos de menor tamaño de muestra<sup>86</sup>. También se han propuesto análisis de regresión lineal de los valores del efecto sobre las puntuaciones o los componentes de calidad cuando hay un número amplio de ensayos clínicos<sup>48</sup>.

#### *Tercer método: análisis de sensibilidad*

El tercer método en la evaluación de la influencia de la calidad metodológica de cada ensayo sobre la revisión sistemática es el análisis de sensibilidad. Algunas veces se realizan revisiones en las que se analizan separadamente los ensayos que reúnen mejores criterios de calidad, por ejemplo, aquellos con puntuaciones por encima de la mediana en la escala, y sus resultados se comparan con los del conjunto de ensayos seleccionados<sup>87</sup>. También se pueden estratificar los ensayos clínicos según sus puntuaciones en la escala de calidad y comparar entre sí los valores del efecto en cada estrato, de modo que, si difieren, se toman los estimadores del estrato superior como los válidos<sup>88</sup>. Se han propuesto otros tipos de análisis, desde comparar el metaanálisis que tiene en cuenta las puntuaciones de calidad con el que no las tiene en cuenta, hasta efectuar múltiples metaanálisis con omisión de un estudio cada vez, pasando por análisis que omiten los estudios más influyentes o los menos significativos o gráficos que referencian los valores del efecto según distintas características de los pacientes<sup>89</sup>. Varios autores proponen un sistema de metaanálisis acumulado partiendo de un análisis sólo con el ensayo de mejor puntuación de calidad y añadiéndolo uno a uno los demás ensayos en orden decreciente de puntuación, con el objeto de obtener una serie de medidores de efecto y de intervalos de confianza progresivos que se relacionan con las puntuaciones mínimas incluidas de la escala de calidad en un “gráfico del impacto secuencial de la calidad” de carácter exploratorio<sup>48,90,91</sup>.

#### *Cuarto método: ponderación de las escalas de calidad en el metaanálisis*

El cuarto método es incorporar las puntuaciones de la escala de calidad de cada ensayo como pesos en el metaanálisis, al igual que lo que suele llevarse a cabo con sus tamaños de muestra o sus variancias<sup>92</sup>. La ponderación plantea dificultades estadísticas, ya que supone que la escala expresa una relación lineal, lo que tiene que demostrarse y su recorrido puede alterar arbitrariamente el intervalo de confianza del valor del efecto<sup>48,93</sup>. Esta opción de la ponderación con las puntuaciones de calidad se ha usado poco<sup>70</sup>.

#### **Evidencia empírica a favor de la evaluación de la validez interna de los ensayos en las revisiones sistemáticas**

¿Qué evidencia empírica hay sobre la utilidad de la inclusión de estas escalas globales de calidad cuando se aplican a las revisiones sistemáticas? En otras palabras, ¿de qué modo las escalas de calidad como tales indican o detectan posibles sesgos que afecten a la validez interna de los ensayos en las revisiones? Aquí también se encuentran resultados contrapuestos. Mientras algunos autores encontraron diferentes efectos terapéuticos si tenían en cuenta la calidad de los ensayos<sup>54,61,94</sup> y otros no los hallaron<sup>95</sup>, todos usaron escalas de validez problemática. En los dos casos en los que se ha analizado la calidad con la escala de Jadad<sup>63</sup>, la inclusión de los ensayos con peores puntuaciones de calidad arrojó resultados más favorables a la terapia experimental<sup>70,96</sup>.

#### **Críticas a las escalas de calidad de los ensayos clínicos**

Algunos autores han llamado la atención acerca de la falta de sentido de intentar condensar en una cifra única de una escala de calidad características que mezclan propiedades objetivas del estudio como, por ejemplo, si la asignación aleatoria fue oculta, con esquemas de evaluación subjetivos que juzgan de qué modo pueden intervenir dichos aspectos, como son la propia elección de los apartados de la escala y sus criterios de ponderación<sup>97,98</sup>. Estos autores se han centrado en los metaanálisis de estudios observacionales, pero sus críticas son también aplicables a los ensayos clínicos con distribución aleatoria. Rechazan cualquier escala conjunta, que creen que son inútiles en el mejor de los casos y en el peor engañosas y, en su lugar, propugnan la estratificación o el análisis de regresión de los

resultados del metaanálisis según los distintos componentes individuales que intervienen en la calidad de los ensayos<sup>99,100</sup>. No debe olvidarse que, al no haber un “patrón oro” que indique la validez interna real de un ensayo, el proceso de validación de cualquier escala que pretenda medirla será incompleto, por muchas buenas cualidades psicométricas que tenga. Siempre será dudoso el grado de correspondencia entre la puntuación de la escala y el grado en el que el ensayo evaluado carece de sesgos<sup>68</sup>. En el fondo, el metaanálisis no puede limitarse a ser una técnica para resumir en un valor común resultados de estudios que en muchas ocasiones pueden ser heterogéneos en muchos aspectos, no sólo en los que se juzgan en la evaluación de su calidad sino en otros aspectos distintos como sus criterios de inclusión, el tipo de intervención o sus variables de resultado. Como toda técnica estadística, el metaanálisis debe tener un doble objetivo: resumir resultados dispersos en medidas comunes y reconocer los patrones que presentan los datos. Por desgracia, la mayor parte de los metaanálisis publicados responden al primer objetivo y apenas atienden a la exploración de la diversidad o heterogeneidad presente entre los distintos estudios individuales y de sus posibles razones. Antes que la medición de la calidad en un ejercicio intelectual sistematizador cuando menos arriesgado, los esfuerzos se deben concentrar en la mejora de la calidad de los ensayos individuales<sup>101</sup>.

En conclusión, de toda esta controversia se deduce que existen razones poderosas para evaluar la calidad de los ensayos clínicos antes de poder interpretar adecuadamente sus resultados, pero no hay acuerdo sobre el mejor modo de hacerlo y sí existen suficientes motivos para dudar de que se elaboren escalas aceptadas por todos. Mientras tanto, los componentes individuales de la calidad ofrecen un medio más seguro para juzgar los ensayos. En cuanto a las revisiones sistemáticas, parece que la forma más prudente para evaluar la calidad de los ensayos individuales cuando se efectúa el análisis global es mediante análisis de sensibilidad o de estratificación según los distintos componentes individuales. Sin embargo, al margen de las revisiones sistemáticas, las escalas de calidad, si están validadas en un grado razonable, pueden servir, en determinadas circunstancias, para evaluar grupos de ensayos, como por ejemplo, para estudiar el estado de la cuestión en los ensayos financiados o publicados, las tendencias temporales en su diseño o publicación o el grado de difusión de aspectos decisivos en su diseño.

## **Calidad de la publicación del ensayo clínico**

Esto lleva directamente al terreno de la calidad de la publicación de los ensayos clínicos. Independientemente de cómo se evalúe y se tenga en cuenta la calidad, nadie discute que las publicaciones que describen los ensayos clínicos deben ser transparentes y reflejar adecuadamente sus aspectos más relevantes. Sin una descripción adecuada de los ensayos la evaluación de su calidad se convierte en una tarea imposible<sup>12</sup>. Las investigaciones realizadas han encontrado graves deficiencias en la calidad de la publicación, tanto por omisión en la descripción como por errores en el análisis<sup>14,33,40,102-106</sup>. En los últimos años se han desarrollado varias iniciativas dirigidas a establecer patrones normativos sobre la publicación de la investigación clínica, iniciados en 1987 con los resúmenes de los artículos clínicos<sup>107</sup> y seguidos con dos iniciativas paralelas para normalizar la descripción de los ensayos clínicos<sup>108,109</sup> que, tras unos pasos tanteadores<sup>110,111</sup>, han convergido en la declaración Consolidated Standards of Reporting Trial (CONSORT)<sup>112</sup>. Esta declaración establece unas recomendaciones para la publicación de los ensayos clínicos, recogidas en 21 apartados y una figura agrupados en la estructura tradicional *IMRAD* de un artículo médico original (introducción, material y métodos, resultados y discusión), que van destinadas a conocer mejor su validez interna y externa (véanse anexo I y fig. 1 del artículo de Urrutia et al en este número; págs. 107 y 108). La adopción de estas normas no sólo contribuirá a conocer mejor los aspectos de los ensayos que tienen que ver con su validez, sino que también ayudará a esclarecer la importancia de otros aspectos del diseño cuya relación con la validez interna no es conocida suficientemente en el momento actual.

Aunque la acogida a la iniciativa CONSORT ha sido calurosa en muchas revistas médicas<sup>113-115</sup>, hasta el punto de que ya son 70 las revistas médicas que han declarado haberse sumado a las recomendaciones<sup>116</sup>, la declaración también ha recibido críticas por incluir apartados con escasa evidencia empírica o cuando menos conflictivos<sup>117-120</sup>, por no atenerse al principio del análisis por asignación de tratamientos de un modo riguroso e incurrir en varias inconsistencias<sup>121</sup> y se han señalado problemas prácticos para su aplicación<sup>122</sup>. Indudablemente, la iniciativa debe mejorarse y ponerse a prueba en la práctica. Queda por saber si en realidad servirá para lograr su propósito de mejorar y clarificar la calidad de la publicación<sup>123</sup>.

## Propuestas para el futuro

Otra solución propuesta para minimizar otros inconvenientes, como los sesgos de publicación, pero escasamente aplicada, es la creación de registros obligatorios de ensayos clínicos desde su aprobación por los comités éticos de investigación clínica o por los organismos que la concedan<sup>124-126</sup>, como propone la Colaboración Cochrane y ya es efectiva en Australia desde 1985. En España, en los años ochenta se constituyó por razones administrativas un registro de ensayos que solicitaban la preceptiva autorización ante la Dirección General de Farmacia<sup>127,128</sup>. Este registro es similar a los disponibles en otros países de nuestro entorno, también ligados a la existencia de un requerimiento legal de notificación o autorización de los ensayos clínicos con medicamentos por parte de las autoridades sanitarias e incluso existe un proyecto de intercambio de información entre los Estados miembros de la Unión Europea para la creación de una base de datos global sobre ensayos clínicos con medicamentos<sup>129</sup>. Estos registros, con las debidas garantías de salvaguarda de la confidencialidad de la información, deberían poder ser utilizables para el rastreo de ensayos clínicos con medicamentos. La ampliación de estos registros para incluir no sólo los ensayos con medicamentos o productos sanitarios sino con todo tipo de tratamientos es un reto que merece ser considerado. En este sentido, la Colaboración Cochrane española está trabajando en una recopilación de todos los ensayos clínicos realizados en España en algún momento para su posible uso posterior en la elaboración de revisiones sistemáticas.

Paralelamente, también en cuanto a los ensayos clínicos con medicamentos, merece la pena destacar el movimiento de la Conferencia Internacional de Armonización de los Requisitos Técnicos para el Registro de Medicamentos, que funciona desde principios de los noventa, con copatrocinio de las Autoridades Sanitarias de EE.UU., Europa, Japón y de las asociaciones de la industria farmacéutica de dichas regiones. Esta iniciativa, habitualmente denominada ICH por las siglas de la International Conference of Harmonization, ha producido ya decenas de documentos conjuntos, que recogen la mayoría de los principios de calidad en el diseño, ejecución y notificación de los resultados de los ensayos clínicos<sup>30,130-132</sup>, a la vez que garantizan a las compañías farmacéuticas que los resultados de los ensayos clínicos realizados con arreglo a dichas normas serán aceptados para su evaluación por parte de las distintas autoridades reguladoras de

medicamentos (Food and Drug Administration en EE.UU., Agencia Europea de Evaluación de Medicamentos y Autoridades Nacionales en los Estados miembros de la Unión Europea, y Ministerio de Sanidad del Japón).

Pese al cumplimiento de todas las normas de calidad exigibles, los ensayos clínicos todavía pueden llevar a conclusiones erróneas si fallan en sus premisas u objetivos, en su ejecución, en el análisis de los resultados o en la rapidez de su difusión<sup>116,133,134</sup>. Finalmente, la calidad de los ensayos clínicos sólo se garantizará cuando los investigadores clínicos participen con los metodólogos y con los patrocinadores en el diseño y ejecución de los ensayos y cuando se divulguen sus procedimientos, métodos y resultados con transparencia, como el desarrollo de la tecnología de la información pone ya a nuestro alcance.

## Agradecimiento

A Teresa Honrubia Fernández, por sus valiosos comentarios a un borrador anterior.

## BIBLIOGRAFÍA

1. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J* 1979; 12: 1.193-1.254.
2. U.S. Preventive Services Task Force. *Guide to Clinical Preventive Services*. US. Department of Health and Human Services. (2.ª ed.). Baltimore: Williams and Wilkins, 1996.
3. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group. User's guides to the medical literature-IX. A method for grading health care recommendations. *JAMA* 1995; 274: 1.800-1.804.
4. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995; 16: 62-73.
5. Last JM. *A dictionary of epidemiology* (2.ª ed.). Nueva York: Oxford University Press, 1988.
6. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72: 233-240.
7. Altman DG. Randomization: essential for reducing bias. *Br Med J* 1991; 302: 1.481-1.482.
8. Meinert CL. *Clinical trials - Design, conduct, and analysis*. Nueva York: Oxford University Press, 1986.
9. Pocock SJ. *Clinical trials - A practical approach*. Chichester: John Wiley & Sons, 1983.
10. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of the therapy. I: Medical. *Stat Med* 1989; 8: 441-454.

11. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309: 1.358-1.361.
12. Schultz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-412.
13. Schultz KF. Subverting randomization in controlled trials. *JAMA* 1995; 274: 1.456-1.458.
14. Gøtzsche P. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10: 31-56 [corrección *Control Clin Trials* 1989; 10: 356].
15. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32: 51-63.
16. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold: a prophylactic and therapeutic trial. *JAMA* 1975; 231: 1.038-1.042.
17. Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgical or nonsurgical tests for transient ischemic attacks and cervical carotid artery lesions. *JAMA* 1970; 211: 1.993-2.003.
18. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology - A basic science for clinical medicine* (2ª ed.). Boston: Little Brown and company, 1991.
19. Meinert CL. *Clinical trials dictionary. Terminology and usage recommendations*. Baltimore: The Johns Hopkins Center for Clinical Trials, 1996.
20. The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med* 1980; 303: 1.038-1.041.
21. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992; 21: 837-841.
22. Godfrey K. Comparing the means of several groups. *N Engl J Med* 1985; 313: 1.450-1.456.
23. Bulpitt CJ. Subgroup analysis. *Lancet* 1988; 2: 31-34.
24. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991; 266: 93-98.
25. Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. *Ann Intern Med* 1992; 116: 76-84.
26. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; 301: 1.410-1.412.
27. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* 1992; 327: 227-233.
28. Epstein AE, Hallstrom AP, Rogers WJ, Liebson PR, Seals AA, Anderson FL et al. Mortality following ventricular arrhythmia suppression by encainide, flecainide and moricizine after myocardial infarction. The original design concept of the Cardiac Arrhythmia Suppression Trial (CAST). *JAMA* 1993; 270: 2.451-2.455.
29. Khan KS, Daya S, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertil Steril* 1996; 65: 939-945.
30. International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use: Good clinical practice. CPMP/135/95. <http://www.eudra.org/humandocs/humans/ICH.htm> [visualizado el 7 de octubre de 1998].
31. Weiss RB, Vogelzang NJ, Peterson BA, Panasci LC, Carpenter JT, Gavigan M et al. A successful system of scientific data audits for clinical trials. A report from the Cancer and Leukemia Group B. *JAMA* 1993; 270: 459-464.
32. Cohen J. Clinical trial monitoring: hit or miss? *Science* 1994; 264: 1.534-1.537.
33. García-López FJ. Impacto de la normativa legal en la calidad de los ensayos clínicos realizados en España. *Med Clin (Barc)* 1993; 100: 770-777.
34. Rochon PA, Gurwitz JH, Cheung M, Hayes JA, Chalmers TC. Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal. *JAMA* 1994; 272: 108-113.
35. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Tech Assess Health Care* 1996; 12: 195-208.
36. Badgley RF. An assessment of research methods reported in 103 scientific articles from two Canadian medical journals. *Can Med Assoc J* 1961; 85: 246-250.
37. Bland JM, Jones DR, Bennett S, Cook DG, Haines AP, MacFarlane AJ. Is the clinical trial evidence about new drugs statistically adequate? *Br J Clin Pharmacol* 1985; 19: 155-160.
38. Thomson ME, Kramer MS. Methodological standards for controlled clinical trials of early contact and maternal-infant behavior. *Pediatrics* 1984; 73: 294-300.
39. Weintraub M. How to critically assess clinical drug trials. *Drug Ther* 1982; 12: 131-148.
40. DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1.332-1.337.
41. Grant A. Reporting clinical trials. *Br J Obstet Gynaecol* 1989; 96: 397-400.
42. Mahon WA, Daniel EE. A method for the assessment or reports of drug trials. *Can Med Assoc J* 1964; 90: 565-569.
43. Gardner MJ, Machin D, Campbell MJ. Use of check lists in assessing the statistical content of medical studies. En: *Statistics with confidence: confidence intervals and statistical guidelines*. Londres: BMJ 1989; 101-108.
44. Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970; 3: 637-640.



45. Sacristán JA, Soto J, Galende I. Evaluación crítica de ensayos clínicos. *Med Clin (Barc)* 1993; 100: 780-787.
46. Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2: 31-49.
47. Chalmers I, Adams M, Dickersin K, Hetherington J, Tarnow-Mordi W, Meinert C et al. A cohort study of summary reports of controlled trials. *JAMA* 1990; 263: 1.401-1.405.
48. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; 45: 225-265.
49. Poynard T. Evaluation de la qualité méthodologique des essais thérapeutiques randomisés. *La Presse Medicale* 1988; 17: 315-318.
50. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989; 84: 815-827.
51. Spitzer WO, Lawrence V, Dales R, Hill G, Archer MC, Clark P et al. Links between passive smoking and disease: a best evidence synthesis. *Clin Invest Med* 1990; 13: 17-42.
52. Andrew E. Method for assessment of the reporting standard of clinical trials with roentgen contrast media. *Acta Radiol Diag* 1984; 25: 55-58.
53. Brown SA. Measurement of quality of primary studies for meta-analysis. *Nursing Res* 1991; 40: 352-355.
54. Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vendenbroucke JP et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992; 340: 152-156.
55. Ongenhenia P, Van Houdenhove B. Anti-depressants-induced analgesia in chronic non-malignant pain: a meta-analysis of 39 placebo-controlled studies. *Pain* 1992; 49: 205-219.
56. Smith K, Cook D, Guyatt GH, Madhavan J, Oxman AD. Respiratory muscle training in chronic airflow limitations: a meta-analysis. *Am Rev Resp Dis* 1992; 145: 533-539.
57. Ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990; 43: 1.191-1.199.
58. Cho MK, Bero LA. Instruments for assessing the quality of drugs studies published in the medical literature. *JAMA* 1994; 272: 101-104.
59. Goodman SN, Berlin J, Fletcher RH, Fletcher SW. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann Intern Med* 1994; 121: 11-21.
60. Evans M, Pollock AV. A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection. *Br J Surg* 1985; 72: 256-260.
61. Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials. *Ann Intern Med* 1990; 113: 299-307.
62. Imperiale TF, Petrusis AS. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. *JAMA* 1991; 266: 260-265.
63. Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds JM, Gavaghan DJ et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17: 1-12.
64. Kleijnen J, Knipschild P, Ter Riet G. Clinical trials of homeopathy. *Br Med J* 1991; 302: 316-323.
65. Koes BW, Assendelft WJJ, Van der Heijden GJMG, Bouter LM, Knipschild PG. Spinal manipulation and mobilization for back and neck pain: a blinded review. *Br Med J* 1991; 303: 1.298-1.303.
66. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use (2ª ed.). Oxford: Oxford Medical Publications, 1995.
67. Bero L, Rennie D. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; 274: 1.935-1.938.
68. Mulrow CD, Oxman AD, editores. Critical appraisal of studies. *Cochrane Collaboration Handbook* [actualizado el 1 marzo 1997]. Section 6. En: *The Cochrane Library* [database on disk and CDROM]. The Cochrane Collaboration. Oxford: Update Software, 1996 [actualizado trimestralmente].
69. Fahey T, Stocks N, Thomas T. Quantitative systematic review of randomised controlled trials comparing antibiotic with placebo for acute cough in adults. *Br Med J* 1998; 316: 906-910.
70. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M et al. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352: 609-613.
71. Powe NR, Kinnison ML, Steinberg EP. Quality assessment of randomized controlled trials of contrast media. *Radiology* 1989; 170: 377-380.
72. Andrew E, Eide H, Fuglerud P, Hagen EK, Kristofersen DT, Lambrechts M et al. Publications on clinical trials with x-ray contrast media: differences in quality between journals and decades. *Eur J Radiol* 1990; 10: 92-97.
73. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Postdam Consultation on meta-analysis. *J Clin Epidemiol* 1995; 48: 167-171.
74. Delgado-Rodríguez M, Sillero Arenas M. Inclusión de la calidad de una investigación en el metaanálisis. *Gac Sanit* 1995; 9: 265-272.
75. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991; 44: 127-139.
76. Jenicek M. Meta-analysis in medicine: where we are and where we want to go. *J Clin Epidemiol* 1989; 42: 35-44.

77. Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995; 48: 9-18.
78. Spitzer WO. The challenge of meta-analysis. *J Clin Epidemiol* 1995; 48: 1-4.
79. Dickersin K, Chan S, Chalmers TC, Sack HS, Smith H. Publication bias and clinical trials. *Control Clin Trials* 1987; 8: 343-353.
80. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Statist Soc A* 1988; 151: 445-463.
81. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989; 81: 107-115.
82. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1.385-1.389.
83. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-872.
84. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992; 267: 374-378.
85. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A et al. Should unpublished data be included in meta-analysis? Current convictions and controversies. *JAMA* 1993; 269: 2.749-2.753.
86. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997; 315: 629-634.
87. Lam W, Sze PC, Sacks HS, Chalmers TC. Meta-analysis of randomised controlled trials of nicotine chewing gum. *Lancet* 1987; 2: 27-30.
88. Pettiti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis. Methods for quantitative synthesis in medicine. Nueva York: Oxford University Press, 1994; 84-89.
89. Olkin I. Invited commentary: Re: "A critical look at some popular meta-analytic methods". *Am J Epidemiol* 1994; 140: 297-299.
90. Lau J, Antman EM, Jiménez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; 327: 248-254.
91. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol* 1995; 48: 45-57.
92. Klein S, Simes J, Blackburn GL. Total parenteral nutrition and cancer clinical trials. *Cancer* 1986; 58: 1.378-1.386.
93. Basinski A, Naylor CD. Aspirin and fibrinolysis in acute myocardial infarction: meta-analytic evidence for synergy. *J Clin Epidemiol* 1991; 44: 1.085-1.086.
94. Brown AS. Meta-analysis of diabetes patient education research: variations in intervention effects across studies. *Res Nurs Health* 1992; 15: 409-412.
95. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials* 1990; 11: 339-352.
96. Khan KS, Daya S, Jadad AR. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med* 1996; 156: 661-666.
97. Greenland S. A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; 140: 290-296.
98. Shapiro S. Meta-analysis/Shemeta-analysis. *Am J Epidemiol* 1994; 140: 771-778.
99. Greenland S. Reply to "Re: A critical look at some popular analytic methods". Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; 140: 300-301.
100. Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; 351: 123-127.
101. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995; 48: 71-79.
102. 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: 426-432.
103. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. *Br Med J* 1990; 263: 1.355-1.357.
104. 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.
105. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272: 122-124.
106. Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J et al. Completeness of reporting in trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; 347: 363-366.
107. Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; 106: 598-604.
108. The Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Checklist of information for inclusion in reports of clinical trials. *Ann Intern Med* 1996; 124: 741-743.
109. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272: 1.926-1.931.
110. Williams JW, Holleman DR, Samsa FP, Simel DL. Randomized controlled trial of 3 vs 10 day of trimethoprim/sulfamethoxazole for acute maxillary sinusitis. *JAMA* 1995; 273: 1.015-1.021.
111. Rennie D. Reporting randomized controlled trials: an experiment and a call for responses from readers. *JAMA* 1995; 273: 1.054-1.055.

112. 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: 637-639.
113. Rennie D. How to report randomized controlled trials: the CONSORT statement. *JAMA* 1996; 276: 649.
114. McNamee D, Horton R. Lies, damn lies, and reports of RCTs. *Lancet* 1996; 348: 562.
115. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *Br Med J* 1996; 313: 570-571.
116. Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. *JAMA* 1998; 351: 319-320.
117. Black N. CONSORT. *Lancet* 1996; 348: 756-757.
118. Horton R. CONSORT. *Lancet* 1996; 348: 757.
119. Peto R. Clinical trial reporting. *Lancet* 1996; 348: 894-895.
120. Meade TW, Wald N, Collins R. CONSORT statement on the reporting standards of clinical trials. Recommendations are inappropriate for the reports. *Br Med J* 1997; 314: 1.126.
121. Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. *JAMA* 1998; 279: 1.487-1.489.
122. Junker C, Egger M, Schneider M, Zellweger T, Antes G. The CONSORT statement. *JAMA* 1996; 276: 1.876-1.877.
123. Mother D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. *JAMA* 1998; 279: 1.489-1.491.
124. Meinert CL. Toward prospective registration of clinical trials. *Control Clin Trials* 1988; 9: 1-5.
125. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *Br Med J* 1997; 315: 617-619.
126. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986; 4: 1.539-1.541.
127. Ensayos clínicos en España (1982-1988). Madrid: Ministerio de Sanidad y Consumo, Monografías Técnicas, n° 17, 1990.
128. Dickersin K, García-López F. Regulatory process effects clinical trial registration in Spain. *Control Clin Trials* 1992; 13: 507-512.
129. Propuesta de directiva del Parlamento Europeo y del Consejo relativa a la aproximación de las disposiciones legislativas, reglamentarias y administrativas sobre la aplicación de buenas prácticas clínicas en la realización de ensayos clínicos con medicamentos de uso humano. COM (97) 369 final. Luxemburgo: Oficina de Publicaciones Oficiales de las Comunidades Europeas, 1997 (CB-CO-97-451-ES-C).
130. International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use: Structure and content of clinical study reports. CPMP/137/95. [Http://www.eudra.org/humandocs/humans/ICH.htm](http://www.eudra.org/humandocs/humans/ICH.htm) [visualizado el 7 de octubre de 1998].
131. International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Humans Use: General Considerations for clinical trials. CPMP/291/96. [Http://www.eudra.org/humandocs/humans/ICH.htm](http://www.eudra.org/humandocs/humans/ICH.htm) [visualizado el 7 de octubre de 1998].
132. International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Humans Use: Statistical principles for clinical trials. CPMP/363/96. [Http://www.eudra.org/humandocs/humans/ICH.htm](http://www.eudra.org/humandocs/humans/ICH.htm) [visualizado el 7 de octubre de 1998].
133. Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; 279: 281-286.
134. Ioannidis JPA, Lau J. Can quality of clinical trials and meta-analyses be quantified? *Lancet* 1998; 352: 590-591.

Parte 4

---

**ÉTICA MÉDICA**

---

---

# Investigación en humanos

---

J. Moreno<sup>a</sup> y B. González Gabaldón<sup>b</sup>

<sup>a</sup>Comisión de Investigación y Ensayos Clínicos. Hospital Universitario Virgen de la Macarena. Sevilla.

<sup>b</sup>Departamento de Psicología Social. Facultad de Psicología. Sevilla.

La experimentación biomédica con seres humanos se abre paso en nuestro contexto profesional, como un tema especialmente delicado y que provoca una mayor y creciente preocupación entre los profesionales sanitarios, ocupando en estos momentos un lugar primordial<sup>1</sup> entre las diversas cuestiones que se plantean las Ciencias de la Salud y la Tecnología en general.

Vivimos hoy en un mundo de controversias<sup>2</sup> donde todos los valores sociales se enjuician, cuestiones y conceptos, que anteriormente se consideraban irrefutables y se sustentaban en normas morales o religiosas, son actualmente motivo de discusión y revisión.

Las Ciencias de la Salud, que en general tienden cada vez más a desarrollarse desde una perspectiva comunitaria, defienden un esquema de valores implícitos (calidad de vida, justicia social distributiva de recursos) muy acordes con los de la sociedad postindustrial, orientados a metas y objetivos sociales relacionados con las necesidades y prioridades del contexto social; y de la misma manera reconocen la importancia de tener presente las opciones<sup>3</sup> que los recursos reales puedan hacer factibles.

El respeto hacia el contexto social y hacia los valores que éste detenta y la repercusión que las investigaciones aplicadas puedan tener, es una preocupación común entre los investigadores de la Salud y de los problemas sociales. Así, Murray<sup>4</sup> llama la atención sobre la necesidad de estudiar en su contexto real las posibles implicaciones que puedan tener las investigaciones en general, y sobre todo las investigaciones con humanos: temas como el requerimiento del consentimiento explícito por parte de los individuos, el respeto al derecho a la privacidad (intimidad), a la confidencialidad y a la justicia, han de ser cuestiones ineludibles y prioritarias de todo quehacer científico<sup>5-9</sup>.

Si argüimos la necesidad de aceptar, o tener presente, un punto de vista ético, o de justicia social<sup>11-13</sup>, que guíe las investigaciones aplicadas, hemos de tener en cuenta, tanto el sistema ético deontológico (propio de cada profesión), como el sistema ético<sup>6</sup> de las consecuencias de nuestras actuaciones-investigaciones (tabla I). Se trata de tener presente que han de existir unos principios éticos, en las investigaciones con humanos y unos principios éticos<sup>2,5,10</sup> en las aplicaciones e intervenciones<sup>4</sup>.

En esta línea, declaraciones como la de Helsinki y sus continuas revisiones<sup>5</sup>, tienden a plasmar y fijar una serie de principios éticos, a los que deben someterse los investigadores que experimentan con humanos, pero a pesar de ello, continúan presentándose reiterados problemas relativos a la aplicación práctica de esos principios. Una de las mayores dificultades es la imposibilidad de marcar una clara separación entre las investigaciones llamadas básicas o científicas, y las aplicadas o clínicas<sup>14</sup>.

Este hecho es un motivo de continua preocupación para los profesionales de la Salud, para la comunidad científica y para la sociedad en general<sup>1</sup>; para todos ellos se hace evidente la necesidad de llevar a cabo investigaciones con seres humanos<sup>2</sup>, ya que a pesar del buen quehacer científico y de la adecuada metodología en la experimentación con animales, los resultados obtenidos no pueden, en todas las ocasiones, llegar a reproducir, en términos fiables, las respuestas de los seres humanos.

De esta forma en muchas ocasiones se hace insoslayable realizar las investigaciones *in vivo* con seres humanos y es aquí, en estos nuevos estudios biomédicos, donde los profesionales se replantean y reconsideran antiguas concepciones de la vida, de la muerte y del significado del ser.

Los descubrimientos pronostican, según Kieffer<sup>6</sup>, cambios drásticos en nuestro pensamiento

TABLA I  
TEMAS MORALES Y ÉTICOS  
DE LA NUEVA BIOLOGÍA

Experimentación con humanos
Consentimiento libre informado
Líneas maestras para la investigación con humanos: investigación sobre fetos, niños, presos y deficientes mentales
Control social y profesional de la experimentación
Investigación científica
Conciencia científica
Límites a la investigación
Regulación de la libertad de investigación
Organismos y entidades reguladores
La ciencia y el bien público

G.M. Kieffer, *Futures Planning: Biology, Society and Ethical Education. The Science Teacher* 1975; 42(8): 12.

ético y graves implicaciones para la política y las tareas legislativas. Cuestiones tales como ¿tienen los científicos que investigan responsabilidades especiales? obligan a la comunidad a reflexionar sobre las crisis de confianza en la ciencia y sugiere una comprensión más realista de las aplicaciones clínicas y sociales que han surgido y que probablemente surjan.

Muchos de estos temas se extienden más allá de los planteamientos individuales para alcanzar a la comunidad, y de esta forma de preguntas sobre la política social demandan una respuesta comunitaria. Los destinos de la ciencia<sup>15</sup>, la política y la sociedad están entrecruzados inexorablemente y debe establecerse, como hipótesis de trabajo, la cooperación entre los implicados en la aplicación humana de la ciencia y los que hacen la política gubernamental. Esto requiere no sólo un conocimiento adecuado de la ciencia y la tecnología, sino también la comprensión y aceptación de aspectos fundamentales de la existencia humana, “una filosofía que de sentimiento a la vida...”<sup>16</sup>, como exponía a modo de lamento Derek Bok, Rector de la Universidad de Harvard en el artículo que publicó en junio del 88 en el *Harvard Magazine*. Se necesita, pues, una comunidad informada, que juntamente con los científicos, reflexione sobre las implicaciones morales y las causas sociales de las investigaciones de los seres humanos; la aceptación común de los valores que subyacen tras estas investigaciones tiende a aumentar la probabilidad de éxito en las aplicaciones<sup>17</sup>.

## Protección de los derechos humanos

Uno de los argumentos presentados en defensa de los encuestados en crímenes de guerra en los juicios de Nüremberg<sup>5</sup> se basaba en decir que “la adquisición de conocimiento y el avance de los fines sociales deben estar por encima del individuo”, era un claro atentado al concepto de los derechos humanos básicos; la necesidad de proteger los derechos humanos y de clarificar y codificar de formas de comportamiento ético permiten crear el código de Nüremberg, documento en el cual se especifica la relación entre el experimentador y el individuo experimental y cuya validez actual está basada en las propuestas<sup>6</sup> resumidas como sigue: a) consentimiento voluntario del individuo humano; b) resultados fructíferos de la investigación, para la sociedad; c) experimentación animal previa; d) evitación de la lesión innecesaria, tanto física como moral; e) protección del individuo experimental; f) investigadores científicamente cualificados, y g) libertad de interrupción de la investigación por el individuo humano. (The Nüremberg Code); citado por Kieffer<sup>6</sup>, pág 263.

Así, el valor ético primero de la voluntad humana, “lo primero de todo no hacer daño”, estaba salvaguardado. No obstante, la intensificación de las investigaciones en las últimas décadas ha supuesto una enorme disponibilidad de recursos materiales<sup>18</sup> y de fondos económicos destinados a las investigaciones biomédicas con seres humanos (tabla II), aumentando la ingesta de medicamentos prescritos y no prescritos, ello debido fundamentalmente a varias razones: a) aparición de preparados terapéuticos útiles; b) aumento de longevidad, provocando un cambio en la frecuencia de diversas enfermedades, con mayor incidencia de procesos crónicos y degenerativos; c) aumento de la proporción de consultas al médico general resueltas con receta, y d) aumento del consumo de tranquilizantes.

La cantidad de nuevos fármacos<sup>20</sup> aparecidos en los últimos años a mejorado notablemente las posibilidades terapéuticas, pero también ha creado problemas en la misma proporción. En consecuencia, la revolución terapéutica ha impuesto a los médicos, la responsabilidad ineludible de utilizar correctamente los fármacos y ello evidencia un mayor número de experimentos sobre humanos y por número de científicos cada vez mayor, presionados por la necesidad y deseo de publicar, ya que esto da la medida de su actividad investigadora y determina su futuro profesional, lo que aumenta la probabilidad de cometer errores éticos<sup>13</sup>.

Las nuevas y sofisticadas tecnologías terapéuticas para el tratamiento de la enfermedad<sup>19</sup> conllevan inherentemente un potencial de efectos secundarios; aumentados éstos por la dificultad de dilucidar entre lo experimental y lo terapéutico. Deben encontrarse respuestas<sup>9</sup>, por aquellos que toman en serio el problema de la experimentación humana, a tres principales preguntas: *a) ¿qué límites, si los hay, deben imponerse a la investigación y cuáles son sus implicaciones?; b) ¿quién tendría autoridad para formular estos límites?, y c) ¿por qué medios deben imponerse?*

Además, en estos planteamientos, y dificultando aún más su clarificación, entra en juego un conjunto de preferencias, intereses y motivaciones, implícitos o explícitos en las partes implicadas, creando a menudo graves conflictos: el investigador, el ser humano sujeto de la experimentación y el estado que apoya o restringe la investigación<sup>21</sup>.

### La metodología actual de las investigaciones con humanos

Numerosos comités y grupos de trabajo han expresado en los últimos años su opinión, afirmándose en el convencimiento de que la política científica ha de orientar su proyección a problemas de salud, prioritarios del propio contexto, y que es necesario llevar a cabo una relevante y racional selección de los estudios que han de ser potenciados y los proyectos que se han de aprobar.

Los objetivos científicos concretos de la Política Científica<sup>22</sup> para Investigación y Desarrollo Farmacéutico en nuestro país, se pueden resumir en tres ideas básicas: en primer lugar, fomentar la investigación farmacéutica en cantidad y calidad; en segundo lugar, corregir las deficiencias de infraestructuras y equipamiento, y por último, potenciar y dirigir la formación personal y equipos de investigación hacia áreas de especial interés sanitario y socioeconómico... Estos conceptos, válidos para la política científica en general, aparecen con mayor claridad en el campo de la investigación biomédica y de la salud comunitaria. Quienes se desenvuelven en el campo de la investigación biomédica deben ser plenamente conscientes de su responsabilidad social, y del control público en la justificación de estas investigaciones, ya que los fondos<sup>12</sup> que se adjudican pueden tener otros usos, que contribuyan igualmente al bienestar de la población; los recursos asignados podrían ser utilizados en otras áreas, también críticas, de la atención de la salud. Este hecho crea, en los investigadores, la necesidad de estar seguros de que sus estu-

TABLA II  
ASPECTOS PUNTALES DE LA  
INVESTIGACIÓN FARMACÉUTICA

---

En EE.UU. en el año 1986 se invirtieron 600.000 millones de pesetas en investigación farmacéutica y en Alemania 250.000 millones de pesetas (Pharmaceutical Manufacturers Association 1986)
Durante el año 1988 el coste del consumo de medicamentos en España ascendió a 517.000 millones de pesetas (Ministerio de Sanidad)
Con el Plan Nacional de Investigación Científica y Desarrollo Tecnológico se espera que el presupuesto nacional español para investigaciones biomédicas, alcance en los próximos años los 250.000 millones de pesetas <sup>19</sup>
Entre 1982 y 1986 se obtuvieron 43 fármacos cardiovasculares, 29 psicotrópicos y 26 antiinflamatorios <sup>19</sup>
El coste de lanzamiento de un nuevo medicamento puede llegar a costar 20.000 millones de pesetas <sup>19</sup>

---

dios contribuirán a mejorar la salud y justificar de esta forma la validez de sus esfuerzos.

Las investigaciones con humanos es un campo especializado y multiprofesional, y resulta complejo establecer una política científica racional; son los propios investigadores quienes deben promover esta responsabilidad. Ello requiere la definición de dos puntos fundamentales y su desarrollo posterior: *a) ¿cuáles son los problemas prioritarios?, y b) ¿cuál debe ser la secuencia de estudios que haga posible las decisiones racionales en cada uno de los problemas?*

Lo que adquiere importancia decisiva son los factores culturales que privan en este campo, y que determinan que los responsables de desarrollar investigaciones con humanos<sup>5</sup> tomen como modelo a la investigación de laboratorio. La duración de las investigaciones, su reproducibilidad, el tipo de información a manejar, la utilización de grupos de control, la calibración de las técnicas de observación utilizadas, la elaboración de estándares para la interpretación de datos así como la evaluación, señalan aspectos de profundas diferencias entre la investigación de laboratorio y las investigaciones clínicas, indicando la necesidad de un enfoque básico de éstas sus aspectos específicos, creando mejores métodos de

observación, identificación y clasificación de pacientes.

La poderosa tecnología terapéutica lograda<sup>19</sup>, no se ha visto seguida por un simultáneo desarrollo científico de los métodos que dirijan la adopción de decisiones, acerca de cuánto, cómo y en qué pacientes aplicar esta tecnología y la consecuencia de este subdesarrollo científico son las controversias masivas que existen hoy en muchos campos de la medicina y la cirugía.

La compleja sociedad de nuestros días exige cada vez más el estudio de problemas multifactoriales que deben ser analizados en su ámbito natural y cuya descomposición en subproblemas, que son estudiados aisladamente, no siempre asegura el conocimiento global del fenómeno.

La determinación de los límites que deben imponerse a la investigación con humanos, ha de encontrar respuesta tanto en el investigador<sup>3</sup>, como en el sujeto de experimentación. Dentro de la búsqueda de preferencias y motivaciones, el consentimiento del individuo, adquiere mayor relevancia. Pero, ¿qué problemas contiene el consentimiento?

La dificultad existente a la hora de proteger los derechos<sup>7</sup> de los pacientes para determinar la información han de recibir, nos obliga a plantear algunas regulaciones como las siguientes: *a)* explicación sincera; *b)* descripción de riesgos e incomodidades; *c)* descripción de beneficios; *d)* exposición de procedimientos; *e)* contestación a cualquier pregunta, y *f)* libertad de interrumpir la participación.

El estudio de los efectos toxicológicos de los fármacos<sup>23</sup>, obliga a la utilización de personas voluntarias sanas, previos estudios *in vitro* y en animales. Esta situación responsabiliza a los investigadores en requerimientos concretos, sobre extensos estudios en animales, con antelación a la investigación y acreditación científica de la competencia de los investigadores acerca de la experimentación propuesta. Una vez contempladas estas premisas, existe una obligada planificación<sup>5,6,14</sup> cuyo razonamiento se asienta en las distintas etapas y fases de que consta la investigación en seres humanos:

*Fase I.* Prueba de seguridad en voluntarios sanos. Se consideran incluidos en esta fase, los estudios de farmacología clínica que corresponden a la primera administración de la sustancia, a ensayar en el individuo, con el fin de obtener los primeros datos de farmacocinética, metabolismo y farmacodinamia y, eventualmente, la tolerancia del preparado y la relación entre dosificación y efectos colaterales. Este estudio se realizará en

un número reducido de voluntarios sanos, con la administración de una sola dosis o de un número reducido de ellas.

*Fase II.* Prueba terapéutica en pacientes. Se consideran en esta fase, los ensayos controlados de farmacología clínica que comprenden la primera administración de la sustancia a un número limitado de enfermos afectados del proceso o procesos patológicos, para los que se presuponga indicada la sustancia en estudio, con el fin de valorar inicialmente los efectos, las acciones terapéuticas, así como la posología y la tolerancia.

*Fase III.* Prueba de seguridad, eficacia y dosis óptimas en pacientes. Se consideran en esta fase, los ensayos tipo clínico realizados en un número suficientemente extenso de enfermos, afectados del proceso o procesos patológicos, para los que se encuentra indicado el medicamento en estudio, con el fin de comparar la eficacia, seguridad y inocuidad del tratamiento nuevo, respecto de un placebo o de otros tratamientos ya conocidos.

*Fase IV.* Pruebas con medicamentos comercializados.

De entre las clases de ensayos clínicos cabe destacar aquellos que puedan ser no controlados, ensayos que no comportan una comparación con el grupo control o testigo o bien controlados, aquellos que comportan una comparación estadísticamente válida, entre los resultados obtenidos en un grupo tratado con la sustancia de experimentación, y otro grupo control o testigo.

La neutralización del efecto placebo<sup>6</sup> y la subjetividad del observador nos obliga a determinar metodologías denominadas ciegas, donde se intenta que el paciente, y el investigador en algún caso, desconozcan el tratamiento que se suministra a los individuos, controlando a través de códigos esta situación, a fin de solventar cualquier anomalía de efectos adversos o abandonos. Este tipo de modalidad presenta actualmente situaciones controvertidas, debido fundamentalmente a la imposibilidad de dar toda la información conveniente a personas expuestas, y si además el tratamiento es un placebo, puede darse el caso de plantearnos la posibilidad de erosionar la confianza de los pacientes. Está claro que la prescripción de placebos hace imposible el consentimiento informado, no obstante Sissela Bok<sup>24</sup> sugiere la consecución de ciertos principios globales: *a)* Los placebos sólo se deberán utilizar después de una cuidadosa diagnosis;



b) no deberán utilizarse placebos activos; c) no se administrarán a individuos que no lo deseen, y d) no deberán administrarse cuando exista un tratamiento indicado.

Las situaciones conflictivas, en nuestro país, por falta de normas específicas sobre los estudios toxicológicos previos<sup>23</sup>, han quedado solventadas en parte, tras nuestro ingreso en la Comunidad Económica Europea<sup>25,26</sup> al ser de aplicación las normas establecidas en la Recomendación Comunitaria 83/571.

Los ensayos clínicos con productos terapéuticos y preparados medicinales, fueron regulados en España por un Real Decreto de 14 de abril de 1978, que más tarde fue desarrollado en Orden Ministerial de 3 de agosto de 1982. Posteriormente, se ha realizado una revisión de los resultados obtenidos en los primeros 5 años de vigencia de esta ley, procediéndose al estudio de 1.331 expedientes de solicitud de ensayo clínico durante el período 1982-1986<sup>14</sup>. Actualmente la nueva ley del medicamento contempla una mejora sustancial al corregir algunas deficiencias teóricas y modificar aspectos que la experiencia ha demostrado difíciles de llevar a la práctica. Por ejemplo: la obligatoriedad<sup>7,14</sup> de concertar un seguro para los pacientes y la homologación de los actuales Comités de ensayos clínicos a los comités éticos, incluyendo en ellos personas ajenas a la profesión sanitaria, dando así un mayor protagonismo, que permite un control más práctico y directo que el control burocrático tenido hasta ahora.

## BIBLIOGRAFÍA

1. Sánchez V. Ciencia: la ciencia plantea demasiados aspectos negativos. *Diario 16*, 5 septiembre 1988.
2. Acuña HR. Ética médica y educación médica. *Bol Of Sanit Panam* 1981; 90: 471-477.
3. Davis DO, Jason LA. Developing a support network for community psychologist. *J Community Psicol*, 1980; 10: 15-33.
4. Murray TH. Ethics, power, and applied social psychology. *Appl Soc Psychol Annual* 1982; 3: 75-95.
5. Howard-Jones N. Human experimental in historical and etical perspectives. *Soc Sci Med* 1982; 16: 1.429.
6. J. Kieffer GH. *Bioética*. Madrid: Editorial Alhambra SA, 1983.
7. Trias Irubies R. Los derechos del enfermo. *JANO* 1988; XXXV: 45-50.
8. French Paul A. Ethical problems in human experimentation. *N Engl J Med* 1985; 273: 687-692.
9. Gesten EL. Social and Community Interventions. *Annual Rev Psycol* 1987; 38: 427.
10. Barber B. The ethics of experimentation with human subjects. *Sci Am* 1976; 234: 25-31.
11. Sabater F. Problemas en torno a un cambio de civilización: el presente de la Ética. *Ed Nuevo Arte Thor* 1988; 299-306.
12. Uriarc Marsal J. La investigación farmacológica al servicio de la salud. *JANO* 1988; XXXV: 71-76.
13. Vidal M. Ética de la actividad científico-técnica. *Moralvia* 1983; 5: 419-443.
14. Ministerio de Sanidad y Consumo. Dirección General de Farmacia y Medicamentos. *Ensayos Clínicos en España*. Monografía 7. Madrid, 1988.
15. Barriga S. *Psicología del grupo y cambio social*. Barcelona: Editorial Hora, 1983.
16. Asenjo J. Artículo sobre el Rector de la Universidad de Harvard. *ABC* 20 febrero 1989; 17.
17. Silverman. Fundamental role characteristics of the community psychologist. *Commun Psycol* 1978; 6: 207-215.
18. Rovira Forns J. Ética y asignación de recursos en el mercado de servicios médicos. Libro de Ponencias. Granada. Escuela Andaluza de Salud Pública, 1988: 253-260.
19. Ariza LM, Coperias EM. La medicina increíble. *Revista GEO*. 1989 febrero; 11-13.
20. López Iglesias J. Investigación Biomédica en España. *JANO* 1985; 7-24.
21. González Rodríguez B. Medicina y Poder. *JANO* 1985; 652: 71-79.
22. Comisión Internacional de Ciencia y Tecnología. Los programas que componen el PLAN 1988-1991. Madrid, Política Científica, 1988.
23. López Azpitarte E. Psicofarmacología y control de la persona. Aspectos Éticos. *JANO* 1986; 30: 18-23.
24. Sissela Bok. The ethics of giving placebos. *Scientific American* 231, 5, 1974; 17-23.
25. Diario Oficial de las Comunidades Europeas. Recomendaciones del Consejo de 9 de febrero de 1987, relativa a las pruebas para la comercialización de las especialidades farmacéuticas. 16.3.87, n.º L73/1, 1987.
26. Olburghs. Documento de sesión sobre materia ética en medicina. PE 106.538/def. 2WG (U8A)/4287S. Comunidades Europeas 1988-1989.

---

# Entre el ensayo clínico y la ética clínica: las buenas prácticas clínicas

---

Diego Gracia

Departamento de Salud Pública e Historia de la Ciencia. Complutense de Madrid.

El término “buenas prácticas clínicas” (BPC) fue introducido en el idioma inglés por la industria farmacéutica norteamericana para designar las normas que en los años 1977<sup>1</sup> y 1988<sup>2</sup> había comenzado a publicar la FDA con los requisitos administrativos a que debían atenerse los promotores, monitores e investigadores en la realización de los ensayos clínicos (EC). El término se construyó a imagen del ya entonces existente de “buenas prácticas industriales” (BPI), que son las utilizadas en la producción industrial de los medicamentos. Este carácter, de algún modo azaroso de la creación del término BPC, explica que algunos lo consideren confuso e impropio<sup>3</sup>.

En nuestro país la preocupación por las BPC procede de estos últimos años<sup>4,5</sup>. La Comunidad Europea publicó en 1990 sus BPC en forma de recomendaciones que debían asumir en su legislación los países miembros<sup>6,7</sup>. Ese mismo año, la Ley del Medicamento española encomendaba a las administraciones sanitarias, en su artículo 65, punto 7, el velar “por el cumplimiento de las normas de BPC”<sup>8</sup>. Y en el anteproyecto del Real Decreto que desarrolla la Ley de Sanidad, en el punto relativo a los EC, se dedica un artículo, el 19, a las “Normas de Buena Práctica Clínica”<sup>9</sup>.

Las normas de BPC han venido exigidas por los propios hechos. La realización de EC sin los debidos controles de calidad no sólo invalida sus resultados, sino que además deteriora gravemente la asistencia sanitaria. De ahí que las BPC no deban verse como una imposición administrativa y burocrática más, que encarece y dificulta la investigación clínica, sino como unas normas encaminadas a elevar la calidad tanto de la asistencia sanitaria como de la investigación clínica. Las BPC deben considerarse como el punto de confluencia de la asistencia, la investigación y la ética clínica. Si se tiene en cuenta que investigación

y asistencia han ido secularmente por caminos no sólo distintos, sino también con frecuencia antitéticos, cabe decir que nos encontramos en una situación privilegiada, que debemos saludar como el comienzo de una nueva era.

La Ley del Medicamento define el EC como “toda evaluación experimental de una sustancia o medicamento, a través de su administración o aplicación a seres humanos”, con el objeto de probar su eficacia y/o seguridad<sup>8</sup>. Su objetivo es llevar a cabo investigación en la clínica y clínica en la investigación<sup>10</sup>. Esto ha podido provocar grandes resistencias en la sociedad e incluso en los propios médicos, que frecuentemente han considerado como única investigación auténtica la básica o preclínica. Según esta mentalidad, tan ubicua a todo lo largo de la historia de la medicina, lo clínico no es experimental, sino terapéutico, y lo experimental no es clínico, de donde se deduce la primera y quizá más clásica objeción ética al EC: debe considerarse inmoral toda experimentación con seres humanos, y más si éstos se hallan enfermos. El primer objetivo de la medicina es la curación y el cuidado del paciente, no el incremento del saber o la promoción del avance científico. El deber primario del médico es curar, no investigar. En favor de esto suelen aducirse diversos testimonios históricos, cada uno de los cuales parece probar que el olvido de este principio ha llevado siempre a la medicina a auténticas aberraciones. El ejemplo más conocido es el de los experimentos de los campos de concentración nazis, aunque no es el único ni el último<sup>11</sup>. Hay una amplia y respetable literatura sobre este tema<sup>12-14</sup> y alguna otra que no lo es<sup>15</sup>. Este campo es particularmente propicio al sensacionalismo usual de la prensa amarilla.

Pero la tesis de que investigación y asistencia son actividades antitéticas no sólo no es cierta sino que dista mucho de hacer justicia a los hechos, que son bastante más complejos. Desde los orígenes de la medicina griega, la clínica médica

se ha venido definiendo por contraposición a la patología. Los griegos establecieron entre ambas la misma diferencia que en la lógica aristotélica existía entre los universales y los particulares. Por ello consideraron que el objeto de la patología era el estudio de los universales patológicos, las llamadas “especies morbosas” (p. ej., la especie morbosus tuberculosis pulmonar), en tanto que el de la clínica era el estudio de individuos concretos, Pedro o Juan, que padecen, por ejemplo, tuberculosis. La tesis clásica fue que de los universales cabía conocimiento “cierto”, y por lo tanto “ciencia”, mientras que en el orden de los particulares no era posible más que “opinión” y “probabilidad”. Aristóteles lo afirmó claramente, y de él lo tomaron buena parte de la filosofía y la medicina posteriores. Precisamente porque en este orden del conocimiento particular no era posible la certeza sino sólo la opinión (*dóxa*), se consideraba lógico que personas sabias y prudentes defendieran opiniones encontradas (*parà-dóxa*). Todas debían considerarse defendibles y respetables. Éste y no otro es el sentido de la llamada “libertad clínica”: la capacidad que debe tener todo clínico de seguir el procedimiento que “en su opinión” considere el más adecuado en un caso concreto, siempre y cuando esa opinión esté dentro del conjunto de las que una comunidad científica considera correctas y prudentes. Por más que entre esas opiniones unas gocen de mayor aceptación que otras –lo que en la terminología aristotélica se expresaba diciendo que eran “más probables”–, todas las opiniones defendidas por personas sabias y prudentes, aun las menos probables, podían ser elegidas por el médico, a la vista de las circunstancias que concurrían en ese caso concreto. En esto consistía la libertad clínica, en la capacidad de elegir procedimientos poco probables cuando las características peculiares de un caso concreto así lo aconsejaban. Como las decisiones morales son también concretas, esta doctrina se aplicó después al campo de la ética, donde llevó a afirmar que las opiniones menos probables o poco probables, defendidas por moralistas sabios y prudentes, debían considerarse correctas, y por tanto podían ser aplicables a la solución de casos concretos. Este principio ha sido desde el siglo xvii uno de los pilares de la teología moral católica.

Es precisamente en el siglo xvii cuando se inició el proceso de cuantificación de la antigua idea de probabilidad. De la probabilidad cualitativa o subjetiva (lo que opinan las personas sabias y prudentes) se pasa a la probabilidad cuantitativa u objetiva (el cociente entre las respuestas y las posibilidades). La teoría matemática de la probabilidad

nació precisamente para objetivar la clásica idea del conocimiento probable, y asignar un guarismo indicador de la probabilidad a cada opinión o a cada opción. La teoría matemática de la probabilidad nació, pues, para objetivar la probabilidad clínica y la probabilidad moral.

La primera enseñanza de la moderna matemática de la probabilidad fue que las predicciones sólo pueden efectuarse a partir de colectivos relativamente amplios o numerosos. De lo que se deduce que una opinión (p. ej., que un determinado fármaco es eficaz en el tratamiento de una determinada enfermedad) carece de relevancia estadística si no ha sido comprobada en un número suficientemente amplio de casos. Uno de los grandes progresos de la estadística muestras ha sido poner a punto el procedimiento para calcular el tamaño de una muestra en función de la eficacia esperada, de modo que el resultado pueda considerarse significativo<sup>16</sup>. Toda opinión sobre la eficacia de un procedimiento diagnóstico o terapéutico que no haya sido verificada en una muestra significativamente grande carece de relevancia estadística y, por tanto, no alcanza la categoría de “probable”. Su condición es sólo la de “posible”, no la de “probable”. En el orden de los posibles no hay espacio para la libertad clínica o para la libertad moral. Éstas sólo tienen sentido dentro del grupo de las prácticas ya validadas, y siempre que haya razones objetivas para pensar que en un caso concreto una práctica tiene más probabilidad de éxito que otra. Lo que sólo es posible no puede ser considerado procedimiento terapéutico, sino sólo experimental, o mejor, experimentable. La investigación experimental tiene siempre por objeto “ensayar” si el procedimiento posible es, además, probable; y en este caso, cuál es su probabilidad. Tal ensayo, para que sea éticamente correcto, debe cumplir ciertos requisitos mínimos: uno, que haya razones para suponer su eficacia, es decir, que sus beneficios se puedan presumir superiores a sus riesgos; otro, que el diseño sea correcto y los investigadores cualificados; tercero, que los participantes en el ensayo lo sean con conocimiento de causa y voluntariamente, y cuarto, que no pertenezcan a grupos vulnerables o desprotegidos. Éstos son los puntos principales de la ética del EC, o más genéricamente, de la ética de la validación de productos sanitarios<sup>17</sup>.

En la actualidad, la teoría del EC ha llegado a un momento de clara madurez, tanto técnica<sup>18</sup> como ética<sup>17</sup>, a pesar de lo cual los problemas pendientes siguen siendo muchos. Unos son problemas técnicos. No se ha afirmado que la teoría del EC esté agotada, ni que no vaya a sufrir

cambios importantes en el inmediato futuro<sup>19,20</sup>. De hecho, cada vez son más evidentes las insuficiencias de este método tal como lo tenemos hoy<sup>21,22</sup>. Pero aun así hemos de considerarla como la mejor metodología disponible y, por tanto, tenemos razones poderosas para considerar incorrecto cualquier otro procedimiento alternativo. Esto tiene a su vez repercusiones éticas importantes, ya que todo lo técnico y metodológicamente incorrecto debe considerarse, por definición, malo, aunque no todo lo correcto sea sin más bueno. La corrección del EC es una condición necesaria, pero no suficiente. Además de técnica y metodológicamente correctos, los EC tienen que ser éticamente buenos, y para ello han de cumplir con los requisitos exigidos por los cuatro principios de la bioética, el de no maleficencia, el de justicia, el de autonomía y el de beneficencia<sup>23,24</sup>.

A pesar de todos sus problemas, la actual teoría del ensayo clínico no sólo es la más rigurosa que haya existido nunca<sup>25</sup>, sino que además está siendo la punta de lanza de una revolución lógica y ética en medicina de incalculables consecuencias<sup>26</sup>. Lo que está en juego es el propio concepto de clínica. Si en otro tiempo pudo pensarse que la clínica dependía de la "intuición" o el "ojo" del médico, por tanto, de cualidades personales e intransferibles (o si se prefiere, no validables), hoy sabemos que sólo lo validado, o al menos validable, merece el nombre de procedimiento clínico. Esto excluye del ámbito de la clínica a la mayor parte de las llamadas prácticas paralelas, marginales o paramédicas, y también a bastantes de las hoy consideradas usuales o canónicas en medicina. La práctica clínica ya no puede considerarse buena sólo por la categoría de quien la realiza, sino por la corrección técnica de lo que se aplica y de la ética con que se realiza. Durante muchos siglos fue un dogma de la medicina que un tratamiento correcto era el que prescribía un médico debidamente titulado para ejercer su actividad, y que un tratamiento incorrecto era el que aplicaba alguien sin licencia para ejercer, aunque tal tratamiento fuera exactamente idéntico al anterior. Esto explica la importancia que en toda la deontología médica clásica tuvo la figura del "intrusismo", que llevaba a perseguir a quienes ejercían una actividad terapéutica sin estar autorizados para ello, aunque el tratamiento que prescribieran fuera similar al que administraban los médicos autorizados. Y explica también por qué las llamadas "medicinas alternativas" o "medicinas marginales" han sido tradicionalmente admitidas como correctas cuando quienes las ejercían eran los médicos ti-

tulados, e incorrectas en caso contrario. Actualmente tenemos muchas razones para afirmar que la dicotomía no está entre profesionales titulados y no titulados, sino entre prácticas validadas y no validadas. Las prácticas no validadas sólo pueden utilizarse de modo experimental y las no validadas no pueden considerarse nunca como terapéuticas, por más que así lo crean lecciones enteras de médicos titulados.

Todo lo anterior sirve para definir lo que hoy debe entenderse por *lex artis* o BPC, entendida ésta en sentido amplio. En sentido estricto, dentro de la teoría del EC, por BPC se entiende, como ya quedó dicho, el conjunto de normas administrativas que deben respetar los promotores, investigadores y monitores de los EC, a fin de que sus datos puedan considerarse fiables y sean aceptados por las agencias reguladoras de medicamentos. Sin BPC ningún EC alcanzaría su objetivo. Pero el razonamiento inverso también es cierto, razón por la que cabe afirmar que sólo cuando en la práctica clínica de un servicio se ha introducido la filosofía propia del EC, esa práctica clínica merece hoy día el calificativo de buena. Lejos de verse como un mero trámite burocrático y administrativo, las BPC deben considerarse un poderoso instrumento de mejora de la calidad asistencial, de donde resulta que el proceso es circular, ya que el buen EC mejora la calidad de la asistencia clínica y las BPC son imprescindibles para la existencia de EC de calidad. BPC y EC no son conceptos idénticos, pero sí íntimamente relacionados.

Tanto la actualidad como la importancia del tema de las BPC son evidentes. De ahí la conveniencia de conocer la actitud de los promotores<sup>27</sup> y de los médicos españoles ante las BPC<sup>28,29</sup>. Los resultados de las encuestas realizadas por Rafael Dal-Ré a los médicos españoles no dejan de ser sorprendentes, ya que demuestran que un elevado porcentaje de clínicos están a favor de los EC de calidad, comprobados mediante auditorías procedentes tanto de las autoridades sanitarias como de las compañías promotores. También consideran imprescindible el requisito del consentimiento informado, y afirman el derecho del paciente a rechazar la participación en un EC, sin perder por ello su derecho a todos los demás recursos de la asistencia sanitaria. Por más que las conclusiones de estos trabajos no sean fácilmente extrapolables al conjunto de la población sanitaria española, sí demuestran la gran evolución que se está operando en toda o al menos en una parte importante de los profesionales sanitarios españoles, lo cual tiene hoy una especial importancia, ya que con la entrada en vigor del mer-

cado único farmacéutico las empresas promotores van a ser más exigentes en la elección de los equipos, los centros y los países en que realizar sus EC, buscando los que tengan mejor acreditadas sus BPC<sup>30</sup>. Abogar actualmente por las BPC es optar por la medicina de calidad<sup>31</sup>. Por eso adoptarlas con rigor y sin tardanza es una de las tareas más urgentes que tiene ante sí la sanidad española.

## BIBLIOGRAFÍA

1. Department of Health, Education and Welfare. Clinical Investigations. Proposed establishment of regulations on obligations of sponsors and monitors. Fed Reg 1977; 42: 49.612-49.630.
2. Department of Health, Education and Welfare. Obligations of clinical investigators of regulated articles. Fed Reg 1978; 43: 35.210-35.236.
3. Gerlis L. Good clinical practice in clinical research. Lancet 1989; 1: 1.008-1.009.
4. Dal-Ré Saavedra R. Investigación clínica: el monitor de ensayos clínicos y la "buena práctica clínica". Med Clin (Barc) 1989; 92: 174-177.
5. Vallvé C. Buena práctica clínica: Recomendaciones internacionales en investigación terapéutica. Madrid: Farmaindustria, 1990.
6. CPMP working Party on Efficacy of medical products. Good clinical practice for trials on medical products in the European Community. III/3976/88-EN final (11/7/1990).
7. Elizalde J. Bioethics as a new human rights emphasis in european research policy. Kennedy Institute of Ethics Journal 1992; 2: 159-169.
8. Ley 25/1990, de 20 de diciembre, del Medicamento. BOE 22 de diciembre 1990, 38.237-38.238.
9. Dirección General de Farmacia y Productos Sanitarios. Anteproyecto de Real Decreto por el que se establecen los requisitos para la realización de ensayos clínicos. Borrador de 18 de septiembre de 1991 (mecanografiado).
10. Porta Serra M. Métodos de investigación clínica: errores, falacias y desafíos. Med Clin (Barc) 1990; 94: 107-115.
11. Beecher DJ. Ethics and human experimentation. N Engl J Med 1966; 274: 11.959.
12. Lock S. Scientific misconduct. Br Med J 1988; 297: 753-754.
13. Lock S. Misconduct in medical research: does it exist in Britain? Br Med J 1988; 297: 1.531-1.535.
14. Hilgartner S. Research fraud, misconduct and the IRB. IRB 1990; 12: 1-4.
15. Marzouki D. La experimentación en el hombre. La otra cara de la medicina. Madrid: Júcar, 1982.
16. Carné X, Moreno V, Porta Serra M, Veliilla E. El cálculo del número de pacientes necesarios en la planificación de un estudio clínico. Med Clin (Barc) 1989; 92: 72-77.
17. Levine RJ. Ethics and Regulation of Clinical Research. (2.ª ed.) Baltimore-Munich: Urban and Schwarzenberg, 1986.
18. Meined CL, Tonascia S. Clinical trials: design, conduct and analysis. Nueva York: Oxford University Press, 1986.
19. Kadane JB. Progress toward a more ethical method for clinical trials. J Med Philos 1986; 11: 385-404.
20. Clayton D. Ethically optimized designs. Br J Clin Pharmacol 1982; 13: 469-480.
21. Pérez-Mateo M, Boix V, Martínez A, Latour J. Análisis crítico de los ensayos clínicos en pancreatitis aguda publicados entre 1983 y 1989. Med Clin (Barc) 1991; 96: 371-373.
22. Fernández-Ballart JD, Vobecky J, Martí-Henneberg C. Metaanálisis: síntesis e integración de los resultados de estudios independientes en medicina. Med Clin (Barc) 1991; 96: 382-387.
23. Ensayos Clínicos en España (1982-1988). Madrid: Ministerio de Sanidad y Consumo, 1990; 121-131.
24. Gracia D. El ensayo clínico en oncología: Problemas éticos. Rev Cancer (Madrid) 1990; 4: 30-34.
25. García Alonso F, Bakke OM. Metodología del ensayo clínico. Monografía Dr. Antonio Esteve, n.º 11. Barcelona: Fundación Dr. Antonio Esteve, 1991.
26. Rothman DJ. Strangers at the Bedside: A history of how Law and Bioethics transform medical decision making. Nueva York: Basic Books, 1991.
27. Segú JL, Lahuerta dal Ré J. Normas de buena práctica clínica en los ensayos con medicamentos realizados en España. Encuesta a la industria farmacéutica. Med Clin (Barc) 1992; 99: 106-109.
28. Dal-Ré R. Buena práctica clínica en ensayos clínicos: Las responsabilidades del investigador. Encuesta a 827 médicos hospitalarios (I). Disponibilidad, monitorización, seguridad, contrato. Med Clin (Barc) 1993; 100: 346-350.
29. Dal-Ré R. Buena práctica clínica en ensayos clínicos: las responsabilidades del investigador. 2: Consentimiento del enfermo. Med Clin (Barc) 1993. En prensa.
30. Rubio Terrés C. La regulación de los medicamentos en España. Med Clin (Barc) 1992; 98: 741-748.
31. Silverrnan M, Lydecker M, Lee PhR. Bad medicine: The prescription drug industry in the third world. Starvford, Cal.: Stanford University Press, 1992.

---

# Problemas éticos de la utilización de seres humanos en investigación biomédica. Consecuencias sobre la prescripción de medicamentos

---

Inés Galende Domínguez

Dirección General de Farmacia y Productos Sanitarios. Ministerio de Sanidad y Consumo. Madrid.

Recordando un poco la historia, fue como consecuencia de los horrores de la experimentación nazi durante la Segunda Guerra Mundial, y del proceso judicial que se siguió contra los responsables de este hecho, que surgió el primer código referente a la investigación en seres humanos: el Código de Núremberg de 1947. En él se recoge una serie de reglas o normas generales para guiar a los investigadores en la realización de investigación con seres humanos de forma ética. La investigación debe ser diseñada y estar basada en resultados obtenidos de la experimentación animal, y en el conocimiento de la historia natural de la enfermedad o problema en estudio. El resultado esperado deberá justificar la realización de la investigación<sup>1</sup>.

Posteriormente, la XVIII Asamblea Médica Mundial reunida en Helsinki en 1964 adoptó una serie de recomendaciones para orientar a los médicos en los trabajos de investigación biomédica con sujetos humanos: la Declaración de Helsinki. Este documento ha sido revisado en 1975 (Tokio), 1983 (Venecia) y 1989 (Hong Kong).

La investigación biomédica con seres humanos debe realizarse de acuerdo con los principios científicos universalmente aceptados, y debe basarse en la realización de los experimentos en animales y de laboratorio, así como en el conocimiento de la literatura científica al respecto<sup>2</sup>.

Sin embargo estos códigos son difíciles de interpretar y de aplicar en situaciones complejas. Por ello, el Congreso Norteamericano designó una comisión (National Commission) para que elaborara un documento sobre la protección de los seres humanos objeto de la experimentación biomédica y conductual: el Informe Belmont de 1979. En él se enuncian los principios éticos básicos a partir de los cuales se puedan "formular, criticar e interpretar reglas o normas específicas",

de aplicación práctica en la investigación clínica con seres humanos.

Estos tres principios éticos básicos son el respeto a las personas, beneficencia y justicia<sup>3</sup>.

*El respeto a las personas* conlleva, a su vez, que se trate a las personas como entes autónomos y que aquellas personas con su autonomía disminuida sean objeto de protección. Se considera como persona autónoma la que es capaz de tomar sus propias decisiones.

El principio de *beneficencia* implica que debemos tratar a las personas no sólo respetando sus decisiones, sino protegiéndolas de daños y asegurando su bienestar, lo que implica en primer lugar, no hacer daño y, en segundo, maximizar el posible beneficio y minimizar los posibles riesgos. Esto ha dado lugar a que algunos autores<sup>4</sup> hayan diferenciado un cuarto principio ético básico, el de *no maleficencia*, identificándolo con el principio hipocrático *non nocere*, lo que simplifica y aclara las cosas ya que así como la no maleficencia es exigible (deber *prima facie*), la beneficencia sólo puede ejercerse con el consentimiento de la persona. La *justicia*, tal como se entiende en el Informe Belmont, se refiere sólo a la justicia distributiva en relación con el reparto equitativo de los beneficios y de las cargas para todos los miembros de una clase similar de personas. En el caso de la investigación, además, debe asegurarse que los avances derivados de la misma no van a ser aplicados a las clases más favorecidas sino también a todos los grupos susceptibles de obtener un beneficio de ella.

El Prof. Gracia<sup>5</sup> clasifica estos cuatro principios éticos básicos en dos niveles, lo que facilita enormemente la toma de decisiones en cada caso particular. En un primer nivel se situarían los principios de no maleficencia y justicia, íntimamente ligados entre sí, y en un segundo nivel, los de autonomía y beneficencia. En aquellos casos en los que se produjera un conflicto entre ambos niveles, los del primero tendrían prioridad frente al segundo.

---

Publicado en *Med Clin (Barc)* 1993; 101 (1): 20-23.

Una vez enunciados estos cuatro principios éticos vamos a ir viendo en qué medida están relacionados con la investigación biomédica con seres humanos y su aplicación práctica.

### **Problemas éticos de la utilización de seres humanos en la investigación biomédica**

Partiendo de la premisa “todo lo que no es correcto desde el punto de vista científico es éticamente inaceptable”, aunque bien es cierto que “no todo lo correcto desde el punto de vista científico es aceptable desde el punto de vista ético”, vamos a analizar los puntos principales en los que la investigación clínica con seres humanos debe respetar cada uno de los principios éticos básicos antes enunciados.

#### *Respeto por las personas: autonomía*

Como ya hemos comentado, el respeto a las personas implica considerar a los sujetos como seres autónomos (capaces de decidir por ellos mismos) y, por tanto, es preciso que antes de ser incluidos en un proyecto de investigación se obtenga su consentimiento.

El *consentimiento informado* es el procedimiento formal para aplicar el principio de autonomía y debe reunir, al menos, tres elementos: voluntariedad, información y comprensión<sup>3,5</sup>.

*Voluntariedad.* Los sujetos deben poder decidir libremente si quieren o no participar en un proyecto de investigación. Ello implica que no existe ningún tipo de presión externa ni influencia indebida en ninguno de sus grados, persuasión, manipulación, ni coerción<sup>6</sup>.

La *persuasión* se ejerce cuando mediante procedimientos racionales se induce a “aceptar libremente” actitudes, valores, intenciones o acciones defendidas por el persuasor. Es muy difícil de evitar en la práctica en la relación médico-enfermo.

La *manipulación* constituye un grado más de presión externa que la persuasión. Consiste en la influencia de forma intencionada y efectiva de una persona sobre otra, alterando las opciones reales o su percepción de elección.

La *coerción* tiene lugar cuando de forma intencionada, se exageran ciertos riesgos o daños indeseados y evitables con el fin de obtener el consentimiento del sujeto.

El carácter voluntario del consentimiento puede ser vulnerado cuando es solicitado por personas en posición de autoridad o de gran influencia sobre el sujeto de la investigación. En estos ca-

sos, debe ser otra persona desprovista de tales vínculos quien lo solicite.

*Información.* Para poder decidir sobre la participación o no en una determinada investigación, cada sujeto debe recibir la información mínima necesaria, adaptada a su nivel de entendimiento, sobre el objetivo, procedimiento del estudio, beneficios esperados y riesgos potenciales, incomodidades derivadas de su participación, posibles alternativas, etc., así como el carácter voluntario de su participación y la posibilidad de retirar su consentimiento, sin perjuicio alguno. Esta información deberá ser dada al sujeto de forma clara y sin prisas, ofreciéndole la oportunidad para consultar posibles dudas o solicitar más información, y dejándole suficiente tiempo para tomar su decisión. ¿Pero, cuánta y cuál es la información mínima? Aunque hay diferencias de criterio son dos las principales posturas al respecto; por un lado los que defienden el llamado “criterio de la práctica profesional” o “del médico razonable”, en cuyo caso es el profesional quien decide el grado de información y, por otro, los defensores del “criterio de la persona razonable”, caso en el que una hipotética persona deseara conocer los riesgos materiales (graves y/o frecuentes)<sup>3</sup>.

*Comprensión.* Para que el consentimiento informado sea válido es preciso, además, que se haya comprendido la información proporcionada. La capacidad de un individuo para comprender (competencia) es función de su inteligencia, habilidad de razonamiento, madurez y lenguaje. Se admite que un sujeto es competente cuando puede tomar sus decisiones según sus conocimientos, escala de valores y metas personales, una vez analizadas las posibles consecuencias de su decisión<sup>7</sup>. Por lo tanto, la competencia de una persona debe estar en relación con algo determinado, y además puede fluctuar en el tiempo.

En el caso de que los sujetos de investigación sean no autónomos o no competentes, se deberá solicitar el *consentimiento por sustitución* a sus representantes, ya ser posible el asentimiento del sujeto. Siempre debe respetarse la decisión del sujeto a no ser que la investigación sea la única forma posible de proporcionarle una terapia a la que de otro modo no tendría acceso.

La *confidencialidad* es una forma de respetar la autonomía de los sujetos, al no revelar datos correspondientes a su persona, a menos que ellos hayan dado su consentimiento. Es definida como la libertad de un individuo para elegir el tiempo y circunstancias bajo las que, y sobre todo, el grado en que sus actitudes, creencias,

conducta y opiniones son comunicadas u ocultadas a otros<sup>6</sup>.

### *Beneficencia*

La beneficencia requiere que procuremos favorecer a los sujetos de la investigación, no exponiéndolos a posibles daños y que la relación beneficio/riesgo, derivada de su participación en el ensayo, sea favorable. Es decir, se deben “maximizar los beneficios previsibles, y minimizar los posibles riesgos e incomodidades”<sup>3,5</sup>. Para ello es preciso que: *a)* se realicen suficientes estudios preclínicos que aseguren un perfil adecuado en cuanto a eficacia y seguridad; *b)* los beneficios esperados sean, al menos, comparables con los de las alternativas disponibles; *c)* se hayan minimizado los posibles riesgos e incomodidades derivados del estudio, excluyendo de la investigación a personas en las que presumiblemente se incrementaría el riesgo, y *d)* se ha previsto un “tratamiento de rescate” en los casos en que proceda. Se ha clasificado el riesgo en distintos grados, lo que facilita el análisis ético de los proyectos de investigación<sup>8</sup>.

*Riesgo menor que el mínimo.* Es aquel habitualmente encontrado en la vida diaria o durante la realización de un examen físico rutinario o de un test psicológico.

*Riesgo mínimo.* Incluye dos tipos de situaciones, la primera es la posibilidad de padecer un trastorno físico o psíquico trivial y la segunda, la posibilidad remota de daño grave o muerte comparable a la que se asume al viajar en avión en un vuelo regular.

*Riesgo mayor que el mínimo.* Es todo aquel no incluido en las categorías anteriores.

Se acepta de forma casi generalizada que la investigación biomédica en seres humanos con beneficio terapéutico para los sujetos participantes deben suponer un riesgo menor que el mínimo, pero son justificables aquellas otras en las que el beneficio esperado es superior a los riesgos que conlleva. Sin embargo, los proyectos de investigación sin beneficio terapéutico para los sujetos, sólo son aceptables si el riesgo es pequeño comparado con el de la propia enfermedad, se espera obtener un conocimiento de gran importancia y no existen otros medios de obtenerlo y, además, el sujeto comprende los motivos y características del estudio y da su consentimiento informado<sup>9</sup>.

### *No maleficencia*

Este principio ético, que junto con el de justicia se sitúa en un nivel superior (nivel 1), obliga a no hacer nada malo a alguien aunque nos lo pida<sup>4</sup>. Por ello, al someter a posibles riesgos, aunque éstos sean mínimos, a los sujetos de una investigación cuando ésta no va a tener validez científica y por tanto va a ser inútil, puede considerarse maleficente. *La validez científica* de un estudio (validez interna) viene determinada por 3 factores fundamentalmente<sup>10</sup>:

*Hipótesis plausible.* Deben existir suficientes datos preclínicos, fisiopatológicos y/o clínicos, que justifiquen la realización de la investigación en seres humanos, en función de la relación beneficio/riesgo esperada.

*Corrección metodológica.* La investigación realizada en humanos con o sin fines terapéuticos, pero de la que se esperan obtener datos importantes y de relevancia clínica debe estar diseñada según el objetivo propuesto, los conocimientos previos, el tipo de investigación, etc. En el caso de la investigación con medicamentos en seres humanos, los estudios que intentan demostrar eficacia (ensayos clínicos) deben cumplir una serie de criterios más o menos estrictos en función de la patología en estudio, alternativas terapéuticas, etc. Sin embargo, los estudios de carácter epidemiológico o de seguridad, que no intentan establecer relaciones causales, sino generar hipótesis, o definir distintos grados de riesgo, suelen ser menos complejos en cuanto a metodología.

Un aspecto, desde el punto de vista ético, que es metodológicamente importante en algunos ensayos clínicos es la utilización de un placebo en el grupo control. Se considera placebo a una sustancia sin actividad farmacológica, que pretenden eliminar los sesgos de valoración cuando se estudian variables subjetivas. La utilización de placebo sólo estaría justificada si se cumplen las siguientes premisas<sup>11</sup>: *a)* no existe tratamiento eficaz para la patología en estudio, o éste presenta un número muy elevado de efectos secundarios importantes, y *b)* la patología en estudio es leve o moderada, y presenta respuesta favorable al placebo.

En ambos casos, siempre deben estar previstas las medidas a tomar en caso que la situación clínica no mejore o empeore.

*Tamaño de la muestra.* Realizar una investigación en un número insuficiente de sujetos para poder extraer conclusiones válidas, es tan



poco ético (tan maleficente) como someter a riesgos innecesarios a un número muy elevado. En el primer caso los riesgos e incomodidades sufridas no habrán servido para nada, y en el segundo caso habría un número de sujetos a los que se estaría privando de la mejor alternativa disponible.

La técnica del metaanálisis intenta, mediante procedimientos estadísticos, obtener conclusiones al tratar conjuntamente los resultados de estudios que por sí solos no pueden proporcionar. Sin embargo, además de ser bastante compleja, es difícil de aplicar ya que requiere criterios homogéneos en los estudios, y ello no es posible en muchas ocasiones.

Los *estudios sin beneficio terapéutico* para los sujetos participantes, en principio, siempre suponen una relación beneficio/riesgo desfavorable, por lo que podrían considerarse como maleficentes. Sin embargo, dado que las únicas investigaciones permitidas son las que no superan el riesgo mínimo, estos estudios pueden aceptarse desde el punto de vista ético, lo que implica que estudios, por ejemplo, con fármacos de elevada toxicidad como los antineoplásicos, no se realicen nunca en voluntarios sanos.

### Justicia

El principio de justicia obliga a tratar a todos los seres humanos con igual consideración y respeto, sin establecer otras diferencias entre ellos que las que redunden en beneficio de todos y en especial de los menos favorecidos<sup>5</sup>. Para ello es necesario distribuir los beneficios y las cargas de la investigación de forma equitativa.

*La selección de los sujetos participantes* mediante unos criterios de inclusión y de exclusión no discriminativos, tanto individual como socialmente. Es decir, no incluyendo personas en función del resultado que queramos obtener (leves o más graves), ni a los que ya llevan más cargas por su posición social<sup>3</sup>.

No se deben incluir *poblaciones vulnerables* (no autónomas y/o no competentes), a no ser que la investigación sólo pueda realizarse en ellas, por estar directamente relacionadas con las condiciones específicas de ese tipo de personas. En este grupo se incluyen las minorías raciales, los económicamente más débiles, y los reclusos en instituciones cerradas. Son poblaciones especialmente vulnerables los niños por lo que también se han redactado normas específicas para realizar investigación en ellos.

En todos los casos será exigible la indemnización por daños derivados de la investigación,

independientemente de que se haya dado el consentimiento para participar en la misma. El principio de justicia también incluye la realización de investigación en países en vías de desarrollo de fármacos que por su elevado coste sólo van a poder ser utilizados en los países desarrollados. Sería un problema de justicia social ya que no habría un reparto equitativo de las cargas y los beneficios, al correr unos con todos los riesgos de la investigación, y si el resultado es favorable, sólo se beneficiarían los otros<sup>12</sup>. No sería aplicable si la patología o condición en estudio fuera específica del área donde se desarrolla la investigación.

### Consecuencias sobre la prescripción de medicamentos

Para que un medicamento pueda ser prescrito es necesario que haya sido autorizada su comercialización por la agencia reguladora correspondiente. Los criterios técnicos en los que se basan para ello, incluyen estudios de eficacia (principalmente ensayos clínicos controlados) y de seguridad, además de considerar las alternativas disponibles para el tratamiento del proceso en que se va a autorizar.

Una vez que el medicamento ya está en el mercado, existen una serie de factores que se deben considerar en relación con su prescripción.

#### Generalización de los resultados

Los ensayos clínicos (EC) que se realizan para demostrar la eficacia de un fármaco en una patología determinada, pueden ser clasificados en dos grandes grupos:

*EC explicativos.* Se caracterizan por incluir a un número reducido de pacientes, con unos criterios de inclusión y exclusión muy estrictos (grado de patología determinado, sin otros procesos patológicos ni tratamientos asociados, etc.) y se realizan en unas condiciones de supervisión y control que no se dan posteriormente en la práctica clínica. Son los que deben hacerse en las primeras fases de investigación.

*EC pragmáticos.* Incluyen a un gran número de pacientes, con criterios más amplios, generalmente son estudios multicéntricos o colaborativos y pretenden asemejarse a las condiciones de uso habituales en la práctica clínica. Estos ensayos, que pueden realizarse tanto pre como poscomercialización, son todavía muy poco frecuentes.

Por tanto, gran parte de los medicamentos antes de su comercialización no han sido probados en condiciones reales de uso, y ello puede explicar que en algunas ocasiones se hayan tenido que retirar del mercado años después, bien por falta de eficacia, o lo que es más frecuente, por producir reacciones adversas graves.

### *Sesgo de publicación*

A la dificultad de extrapolar los resultados obtenidos en la investigación clínica a la práctica se añade otro problema no menos importante como es el sesgo de publicación. Éste puede ser debido a:

*Resultados negativos.* Es bien sabido que se tiende a publicar sólo aquellos estudios en los que los resultados obtenidos son favorables según la opinión del investigador, privando así a la comunidad científica de elementos de juicio para decidir sobre la eficacia/seguridad de un fármaco determinando su posición con relación a las alternativas disponibles<sup>13,14</sup>.

*Manipulación de resultados.* Puede considerarse como uno de los pecados más graves que puede cometer un investigador y, sin embargo, es una práctica relativamente frecuente. En ocasiones eliminando datos que se escapan de un límite prefijado (p. ej., 2-3 desviaciones estándares), o para favorecer a un determinado fármaco. En otras ocasiones, un mismo investigador publica un mismo estudio en 2 revistas distintas, pero con pequeñas modificaciones<sup>5</sup>.

*Resultados confidenciales.* Por su parte, la industria farmacéutica a veces no publica resultados tanto positivos como negativos de estudios realizados con un fármaco de su investigación, hasta pasado cierto tiempo, o hasta que el fármaco está comercializado o incluso, nunca. Se apela casi siempre a la confidencialidad de los datos, resultados preliminares, etc.

*Política editorial.* Las revistas científicas también tienen tendencia a publicar sólo aquellos resultados positivos, que favorecen a un fármaco determinado, o guiadas por una "cierta moda", los resultados de los estudios referidos a una condición o patología dada. A ello debemos añadir la publicación de estudios a los que se exige mayor o menor rigor metodológico y/o ético, y la popularidad o reconocimiento de los autores de la publicación.

### *Prescripción en la práctica clínica*

También en la práctica clínica habitual existen una serie de conductas que podríamos calificar como "poco éticas", en las que están implicados tanto los profesionales médicos como la industria farmacéutica, principalmente. Éstas incluyen los siguientes aspectos.

*Indicaciones no reconocidas.* La prescripción de medicamentos ya comercializados para situaciones o en condiciones de uso distintas de las autorizadas (eficacia/seguridad probadas), muchas veces como respuesta a la publicación de casos clínicos en los que no se han considerado todos los posibles factores implicados en la respuesta obtenida, supone someter a los pacientes a un posible riesgo sin haber demostrado eficacia, y, por tanto, no haber establecido la relación beneficio/riesgo en dicha circunstancia.

*Uso compasivo de medicamentos.* Se entienden como uso compasivo la utilización en pacientes aislados, y al margen de un ensayo clínico, de fármacos experimentales o nuevas indicaciones, en situaciones en las que éste constituye la única alternativa terapéutica. Requiere el consentimiento del paciente y la prescripción médica, no obstante, la compasión no puede autorizar a perpetuar la ignorancia, hacer cosas inútiles ni a dañar al paciente.

*Inducción a la prescripción.* En nuestro medio, la mayor parte de la información científica que recibe el médico prescriptor, y sobre todo el de atención primaria, procede de los delegados de la industria farmacéutica, lo que es utilizado en la mayoría de los casos para inducir a la prescripción del medicamento que les interesa más en cada momento. Esta conducta no sería reprochable si la información proporcionada fuera "aséptica". Sin embargo, lo usual es exagerar deliberadamente las características positivas, omitiendo o enmascarando las negativas.

La inducción a la prescripción también suele hacerse mediante propuestas de "ensayos clínicos promocionales", con los que no se pretende obtener ningún dato nuevo del producto, sino repetirlos incluso sin método alguno, con el único fin de conseguir que el médico prescriptor utilice dicho medicamento.

Es obligación de todos los relacionados de un modo u otro con el mundo del medicamento (médicos e investigadores, industria farmacéutica, autoridades sanitarias y pacientes), exigir rigor tanto metodológico como ético, que como he-

mos visto van estrechamente unidos, porque TODOS somos parte interesada en obtener medicamentos seguros y eficaces.

## BIBLIOGRAFÍA

1. Nüremberg Code 1947. En: *Trials of War Críminals Before the Nüremberg Military Tribunals Under-Control Council Law, vol II, N° 10*. Washington DC: U.S. Government Printing Office, 1949; 181-182.
2. Declaración de Helsinki 1964, Tokio 1975, Venecia 1983, Hong Kong 1989. En: *Ensayos Clínicos en España (1982-1988)*. Monografía Técnica 17. Madrid: Ministerio de Sanidad y Consumo, 1990; 75-78.
3. Informe Belmont 1978: Principios éticos y recomendaciones para la protección de las personas objeto de experimentación. Comisión Nacional para la protección de las personas objeto de experimentación biomédica y conductual. En: *Ensayos Clínicos en España (1982-1988)*. Monografía Técnica 17. Madrid: Ministerio de Sanidad y Consumo, 1990; 121-131.
4. Gracia D. *Primum non nocere*. El principio de no maleficencia como fundamento de la ética médica. Madrid: Real Academia Nacional de Medicina, 1990.
5. Gracia D. *Fundamentos de Bioética*. Madrid: Eudema, 1989.
6. Levine RJ. *Ethics and regulation of clinical research*. (2.ª ed.). New Haven: Yale University Press, 1988.
7. Dyer AR. *Inform Consent Autonomy, and Paternalism: Respecting the patient in Research and Practice*. En: Dyer AR, editor. *Ethics and Psychiatry*. Washington: American Psychiatry Press, 1988; 83-97.
8. Royal College of Physicians. *Guidelines on the practice of Ethics Committees in medical research involving human subjects (2.ª ed.)*. Londres: RCP; 1990.
9. Protección de los sujetos que participan en la investigación. Código de Reglamentación Federal 45 CFR 46. En: *Ensayos Clínicos en España (1982-1988)*. Monografía Técnica 17. Madrid: Ministerio de Sanidad y Consumo, 1990; 159-188.
10. Galende I, Tristán C. Problemas prácticos en un ensayo clínico (II). En: García Alonso F, Bakke OM, editores. *Metodología del ensayo clínico*. Monografías Dr. Antonio Esteve, n.º 11. Barcelona: Fundación Dr. Antonio Esteve, 1991; 11: 21-30.
11. The ACG Committee on FDA-Related Matters. *Placebos in clinical trials of peptic ulcer*. *Am J Gastroenterol* 1989; 84: 469-474.
12. Porter JDH. *The ethics of drug trials in developing countries*. En: Bankowski Z, Bryant JH, Last JM, editores. *Ethics and Epidemiology: International Guidelines*. Proceedings of the XXVth CIOMS Conference. Ginebra: Cloms, 1990; 98-103.
13. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. *Publication bias and clinical trials*. *Controlled Clin Trials* 1987; 8: 343-353.
14. Angell M. *Negative studies*. *N Engl J Med* 1989; 321: 464-466.

---

# Conflicto de intereses e investigación clínica

---

Jordi Camí

Institut Municipal d'Investigació Mèdica (IMIM). Universitat Autònoma de Barcelona.

Las relaciones entre la investigación científica y el sector industrial, afortunadamente, son cada vez más estrechas. Pero la necesaria transferencia de tecnología entre la investigación académica y el sector productivo conlleva un conjunto de riesgos que puede comprometer la integridad de la ciencia. Para algunos el riesgo principal consiste en una excesiva "comercialización" de la propia ciencia, entre cuyas consecuencias destacarían la pérdida de objetividad en el desarrollo y comunicación de los resultados, la reordenación de las prioridades de la investigación hacia programas de mayor aplicabilidad en detrimento de la investigación básica, la degradación de la naturaleza esencial de la ciencia en tanto que empresa abierta, no secreta y colegiada, y la explotación de estudiantes y becarios<sup>1</sup>.

En cuanto a la investigación clínica se refiere, debe reconocerse que la financiación de proyectos científicos depende progresivamente de los recursos procedentes de la industria farmacéutica y sanitaria<sup>2</sup>. Entre este sector y el sanitario se pueden originar situaciones conflictivas donde los intereses creados pueden condicionar la integridad de los resultados y los compromisos de los participantes en la investigación. Se trata de situaciones conocidas con el nombre de conflictos de intereses. Según el Diccionario de la Real Academia de la Lengua Española, un conflicto es una situación desgraciada y de difícil salida. Como se verá a continuación los conflictos de intereses son difíciles de evitar y su existencia perjudica frecuentemente la credibilidad social de la ciencia. El reto consiste en que estas situaciones no comprometan la necesaria promoción de transferencia de tecnología entre la investigación académica y el sector productivo<sup>3</sup>.

La aparición de conflictos de intereses en investigación clínica tiene unos componentes especifi-

cos<sup>3</sup>. Por una parte, la peculiar intervención económica de la industria sanitaria dentro de la sanidad, en tanto que esta industria dispone de mecanismos privilegiados para influir en diversos elementos del mercado, especialmente la demanda. Por otra, la preocupante deshumanización de una asistencia sanitaria que cuenta cada vez con más medios tecnológicos y con menos recursos, todo ello en un entorno sumamente competitivo. En este contexto, algunos profesionales pueden verse fácilmente atraídos por otros intereses, como la fama, el prestigio y la notoriedad o el lucro personal. El problema principal de los conflictos de intereses es que todos pensamos que no seremos nunca objeto de su influencia, y habitualmente no es así.

## Lo que determina los conflictos son las situaciones

---

En palabras de Dennis Thompson, se entiende por conflicto de intereses aquella situación en la que el juicio de un profesional concerniente a su interés primario (p. ej., la salud de los enfermos o la integridad de una investigación) tiende a estar indebidamente influenciado por un interés secundario (de tipo económico o personal)<sup>4</sup>. Con independencia de que el interés secundario sea necesario o deseable, se produce un conflicto cuando éste prevalece sobre el interés primario. Los conflictos de intereses no son directamente el resultado de unos comportamientos, sino unas situaciones en las que se dan determinadas circunstancias. Para el análisis de un conflicto de intereses potencial debe seguirse un procedimiento distinto al de los dilemas bioéticos. Mientras en éstos se trata de escoger entre distintos valores que compiten entre sí por su prioridad, en los conflictos de intereses sólo el interés primario tiene prioridad, ningún otro interés debe dominar al principal.

Existen debates acerca de la necesidad de regular *a priori* estos conflictos. Para algunos las

---

Publicado en *Med Clin (Barc)* 1995; 105 (5): 174-179.

palabras tienen connotaciones peyorativas: la existencia de un conflicto de intereses no indicaría mala práctica *per se* sino que solamente indicaría una situación en la que existirían unos factores predisponentes a un determinado tipo de conducta<sup>5</sup>. Para otros, cuando los intereses existen en potencia, también existe una mayor probabilidad de mala conducta. Es decir, la mera existencia de intereses secundarios exige tanto su declaración previa como su vigilancia, único mecanismo que permite garantizar la integridad del interés primario<sup>4</sup>. Incluso a veces, lamentablemente, estos conflictos van acompañados de verdaderos problemas legales<sup>5</sup>.

Los conflictos de intereses no sólo implican una crisis de lealtad entre los investigadores y su institución, sino que pueden poner en crisis los intereses y objetivos de la propia organización. Un ejemplo de situación conflictiva difícil de manejar la encontramos en ambientes académicos, cuando no quedan perfectamente separados los objetivos comerciales de los educativos. Muchas organizaciones norteamericanas consideran inaceptable la participación de estudiantes y de personal investigador en formación en aquellas investigaciones patrocinadas por la industria. Se parte del principio según el cual las personas en formación deben ser protegidas de aquellos ambientes de trabajo donde no podrán ni discutir ni publicar rápida y públicamente sus resultados de investigación. En nuestro medio no existen criterios establecidos acerca de la participación de estudiantes o personal en formación en investigaciones comerciales. Hasta tal punto que muchos ensayos clínicos en centros sanitarios no se realizarían si no fuera por la colaboración directa de los médicos internos y residentes, personal contratado cuya finalidad principal es formarse en una especialidad médica.

### **“Disclosure of financial interests is the bedrock of any sensible scheme”<sup>6</sup>**

Todos los autores coinciden en que la declaración pública de los intereses existentes es la principal medida para la prevención de conflictos<sup>4,6,7,8</sup>. En EE.UU., se han establecido reglamentos y códigos de conducta desde el ámbito federal (PHS<sup>9</sup> y recientemente los NIH<sup>10</sup>) hasta el ámbito institucional o de centro como en la Escuela de Medicina de Harvard<sup>11</sup>, pasando por las principales asociaciones médicas<sup>12-14</sup>. Aunque su contenido es muy variable entre sí, existe un denominador común entre estos códigos<sup>6</sup>: exigencia de declaración pública de los actuales y potenciales intereses económicos, estricta regu-

lación de la participación financiera o adquisición y venta de acciones en bolsa –sobre todo en el caso de investigadores que participan en ensayos clínicos–, permisividad en la percepción de honorarios o pagos por servicios y establecimiento de comités institucionales para la revisión de las declaraciones. Por ejemplo, en la Escuela de Medicina de Harvard, todo facultativo debe remitir una declaración de intereses anual al comité de integridad de la investigación. Para realizar investigaciones o negocios en los que los profesores o alguno de sus familiares tengan algún interés, debe obtenerse una autorización previa<sup>11</sup>.

En EE.UU. los comités de revisión de declaraciones y problemas sobre la integridad de la investigación tienen ya cierta tradición<sup>1,15</sup>. Nacieron en el ámbito de la investigación sanitaria, gracias a la acción pionera de la Asociación Médica Norteamericana (AMA)<sup>12</sup>, y se han desarrollado después rápidamente en todos los ámbitos de la ciencia académica. Cabe destacar que la responsabilidad de estos comités recae en la propia institución, la instancia responsable de preservar la libertad de actuación académica de sus miembros y, a su vez, de garantizar la autonomía institucional frente a posibles presiones externas<sup>15</sup>. A menudo, estos comités son el último eslabón de todo un proceso de autorregulación interna en el que los acuerdos o situaciones potencialmente conflictivas son consideradas en primer lugar por el responsable superior inmediato, y así sucesivamente, dentro de una sistemática denominada de revisión ascendente<sup>14</sup>.

En España no existe esta tradición, ni en los centros sanitarios ni en la Universidad, centros del CSIC u otros organismos públicos de investigación. Dada la creciente interrelación de la investigación académica con la del sector productivo, los investigadores tenemos la responsabilidad de promover que, tanto en nuestros centros como en nuestros ámbitos asociativos, se empiecen a discutir abiertamente estos temas y nos vayamos dotando de nuestros propios códigos de conducta. Sólo así y de forma complementaria, las instituciones de nuestro país podrán dotarse de comités independientes para canalizar las declaraciones de intereses y velar por la integridad de la investigación.

### **Publicidad de las relaciones entre investigadores y patrocinadores**

La neutralidad de los resultados y conclusiones de una investigación sólo se alcanza tras su exposición pública y su juicio y discusión por parte de terceros. Para ello es necesario que se expon-

ga con absoluta transparencia tanto el motivo como el contexto en el que se ha desarrollado la investigación, es decir, dando una publicidad íntegra de los intereses existentes. La Federación Americana de Investigación Clínica (AFCR)<sup>13</sup> considera que no debería haber excepción alguna a la presentación pública, por escrito u oralmente, de los resultados de cualquier tipo de investigación junto con la declaración de los recursos económicos obtenidos. Aún más, con anterioridad a la realización de los proyectos de investigación no se considera aceptable que exista secreto alguno en el contenido de los acuerdos económicos establecidos. En palabras de Korenman, estos acuerdos deberían ser publicados en las páginas económicas de los periódicos<sup>1</sup>.

En la legítima obtención de compensaciones económicas por parte de los investigadores como contraprestación a sus ideas y trabajos, no importa tanto la cuantía de un determinado servicio sino el tipo de compromisos que se establecen con los promotores<sup>1,6</sup>. Por ejemplo, los resultados de las investigaciones protagonizadas por aquellos “líderes de opinión” que participan reiteradamente en actividades promocionales del sector farmacéutico merecen ser analizados con precaución y cautela<sup>16</sup>. La neutralidad del diseño de una investigación y la interpretación de los resultados obtenidos pueden quedar seriamente comprometidas cuando las compensaciones a obtener están vinculadas al grado de éxito del proyecto. Por ejemplo, en EE.UU., la integridad de algunas investigaciones ha sido puesta en duda tras conocerse que los investigadores habían adquirido acciones en bolsa o habían aceptado participaciones financieras de la empresa interesada<sup>1,17,18</sup>. Llegados a este extremo no es raro que algunos defiendan la inclusión de los pactos económicos alcanzados con el promotor en el texto que se facilita a los enfermos cuando se les invita a participar en un ensayo clínico<sup>19,20</sup>. Con independencia de que haya habido o no una revelación de las relaciones económicas antes del inicio de la investigación, la declaración de intereses se considera inexcusable cuando se publican los resultados. Así lo han acordado y adoptado la mayoría de editores de revistas biomédicas<sup>21</sup>, cuya exigencia fue pionera entre los editores de las revistas científicas de mayor prestigio<sup>22</sup>. Como requisito indispensable para la publicación de los resultados de cualquier investigación clínica, se está generalizando la exigencia a los autores de una declaración previa y por escrito al editor de cualquier posible conflicto de interés<sup>23</sup>. Además, existe el compromiso creciente de los editores de obligar a los autores que añadan al texto de la pu-

blicación la relación de todas las ayudas financieras recibidas. Incluso para la publicación de determinadas investigaciones sobre coste-eficacia, como los estudios de farmacoeconomía, algunos editores ya no consideran suficiente la declaración previa de intereses<sup>24</sup>. y más allá de los conflictos de tipo financiero, es de esperar que las exigencias de declaración previa se extiendan en un futuro no muy lejano hacia otras áreas como la personal, la política, la académica o la religiosa<sup>25</sup>.

Un ejemplo de compromiso exigente lo constituyen las relaciones entre investigadores y promotores del estudio GUSTO, donde se ensayaron varios trombolíticos en el infarto agudo de miocardio. Durante la preparación del ensayo, el comité directivo publicó unas medidas específicas para evitar conflictos de intereses entre el equipo de investigación y las entidades promotoras<sup>26</sup>. Antes del inicio del ensayo clínico, todos los miembros del comité directivo, del centro coordinador y del comité de farmacovigilancia del ensayo GUSTO declararon por escrito que ni ellos ni ninguno de sus familiares inmediatos tenían relación financiera alguna con ninguno de los patrocinadores, incluyendo participaciones financieras, percepción de honorarios, relaciones de consultoría ni reembolso por gastos de viajes. Los investigadores principales de los hospitales colaboradores del ensayo también firmaron una declaración según la cual no tenían participaciones financieras de las compañías patrocinadoras<sup>27</sup>.

### **El desarrollo de nuevos medicamentos, una necesidad con riesgo elevado de conflictos**

Para el desarrollo de nuevos medicamentos, los promotores de ensayos clínicos necesitan atraer con diversos incentivos a los profesionales más competentes y más capaces para este tipo de investigación clínica. Pero, ¿cuáles son las verdaderas razones que motivan la participación de los profesionales en los ensayos clínicos? El médico investigador forma parte de un colectivo cuyos intereses se comparten –a veces con distinta prioridad y a veces en contraposición con los de otros colectivos–. A continuación se describen, a modo de ejemplo, los intereses que pueden tener las distintas partes involucradas en el desarrollo clínico de nuevos medicamentos<sup>28</sup>.

Por una parte está el enfermo, cuyo deseo primario es obtener su curación mediante una asistencia óptima. Por lo tanto, el enfermo esperará que su participación en un ensayo clínico le reporte un beneficio terapéutico adicional; más allá de este beneficio primario, algunos enfermos in-

cluso reconocerán y aceptarán ser partícipes del logro de nuevos y mejores conocimientos científicos. La sociedad en general, como expresión genérica de este colectivo potencial, tendría el mismo orden de intereses; al ser la asistencia sanitaria española un bien común financiado con el pago de nuestros impuestos, la sociedad desea que cualquier intervención en ella, como la realización de ensayos clínicos, redunde en una mejora de la calidad de la asistencia sanitaria.

Por otra parte existen los intereses de los proveedores de asistencia sanitaria que, en España, vienen representados fundamentalmente por el sector público y el privado sin fines de lucro. Si bien los proveedores admiten la realización de ensayos clínicos por razones muy genéricas (oferta de mejoras terapéuticas, fomento inespecífico de la investigación y desarrollo), lo que realmente desean son resultados que les permitan minimizar y racionalizar los costes de la empresa. Los ensayos clínicos aún son un asunto incómodo para algunos responsables de gestionar centros de asistencia sanitaria. Afortunadamente, los centros donde la realización de investigación clínica supone un cúmulo de dificultades tienen cada vez un mayor número de excepciones, ya que se extiende la noción de que la salud es algo más que el alcance de un determinado número de "uvas". Cada vez hay más gestores que valoran los ensayos clínicos y la investigación como un instrumento de prestigio y de publicidad para el centro que dirigen.

El elemento clave es el promotor, sin el cual no existirían la mayoría de ensayos clínicos que se realizan en la actualidad. Su objetivo primario es conseguir el registro de un nuevo medicamento o una nueva indicación, etc., ocupar posteriormente la mayor cuota posible de mercado y obtener los correspondientes beneficios lo antes posible, amortizando la enorme inversión realizada. Secundariamente, la industria farmacéutica también investiga para aumentar su prestigio, ofrecer nuevas o mejores alternativas terapéuticas, mejorar la calidad de la asistencia sanitaria y contribuir al progreso del conocimiento científico.

Por último, los investigadores clínicos. El objetivo primordial de su participación en el desarrollo de nuevos medicamentos es la promoción personal y profesional; secundariamente se persigue un mayor beneficio para los enfermos y se satisface la curiosidad científica. Es decir, se es protagonista de la búsqueda de nuevas y mejores alternativas terapéuticas y se contribuye al conocimiento científico. Los investigadores clínicos realizan ensayos porque obtienen unos legi-

timos incentivos personales y profesionales<sup>17,29</sup>. Entre ellos destaca la obtención de reconocimiento y notoriedad en el sector profesional y la posibilidad de reunirse con colegas, viajar a reuniones internacionales y compartir con líderes científicos de la especialidad. La participación en ensayos clínicos relevantes prestigia al servicio o departamento, de forma que el centro resulta más atractivo para la captación de los mejores residentes y becarios. Otro gran motivo es el reconocimiento que conlleva la publicación de los resultados en revistas científicas de gran difusión, objetivo que no siempre se alcanza al no formar parte de las prioridades del promotor. Por último, y no menos importante, si no hubiera contraprestación económica directa y extraordinaria, el promotor no podría contar con la colaboración de los clínicos para la realización de los ensayos. La financiación constituye un elemento clave para el desarrollo adecuado de la investigación farmacéutica en España y, como veremos a continuación, el tema de discordia cuando aparecen conflictos de intereses<sup>30</sup>.

### La manzana de la discordia

De la realización de un ensayo clínico, naturalmente se derivan unas contraprestaciones económicas. Lógicamente, desde la industria farmacéutica se considera necesario incentivar la motivación de los investigadores clínicos competentes para que participen activa y responsablemente en la realización de ensayos clínicos<sup>29</sup>. Estos incentivos deberán ser implantados y adecuadamente regulados tarde o temprano en nuestro sistema sanitario público, no sin dificultades ya que impera la concepción contractual de médicos con salario homogéneo sin tener en cuenta su productividad y eficiencia. Mientras tanto, no es extraño que se prodiguen prácticas poco transparentes o fronterizas con la mala práctica.

La discusión de los conflictos de intereses se ha centrado principalmente en las consecuencias derivadas de esta contraprestación económica, tanto por el tipo como por su importancia. Pero en España no se puede abordar civilizadamente esta temática sin antes resolver un problema previo, la inexistencia de criterios y normas para la promoción de la investigación en el ámbito clínico. Concretamente, el sistema público de asistencia sanitaria es, a menudo, incapaz de gestionar y administrar adecuadamente los ingresos económicos por ensayos clínicos. En una buena parte de la red sanitaria pública, la financiación estatal y la provisión de servicios sanitarios forma parte de la misma estructura; en otras

palabras, los sistemas financieros son tan centralizados que impiden a los centros cualquier tipo de gestión económica autónoma. Ante esta realidad, se han promovido fundaciones y asociaciones de investigación para la gestión paralela de estos fondos, una solución a medias por las limitaciones y nuevos problemas que ello ocasiona. Pero la limitación más importante es la falta de sensibilidad –traducida luego en incapacidad– de los responsables de los centros sanitarios en todo aquello relacionado con la gestión de los recursos de investigación y de los ensayos clínicos en particular.

Esta situación, en general, tiene consecuencias adversas para todas las partes implicadas. Por ejemplo, muchos ensayos clínicos constituyen un sobrecoste al de la actividad asistencial habitual, ya que el centro no obtiene compensaciones por los gastos extraordinarios derivados del ensayo clínico (fenómeno que puede hacerse extensible a muchas otras investigaciones clínicas). Existen investigadores –cuyo número o proporción nunca se ha cuantificado– a quienes no les queda otro remedio que facturar los emolumentos del ensayo a título particular o en nombre de una sociedad instrumental. Tampoco es raro que algunos investigadores exijan el cobro de unas cantidades para destino personal, incluso en ocasiones, al margen de las que oficialmente pacta el promotor con la entidad, institución o fundación. Si tenemos en cuenta que las compañías farmacéuticas afirman que realizar ensayos clínicos en España no es más económico que en otros países de nuestro entorno, que los gestores de centros sanitarios afirman que los ensayos clínicos suponen un sobrecoste asistencial, y que el coste principal de la asistencia de los enfermos que participan en los ensayos –sea ambulatorio u hospitalario– está subvencionado de forma pública, deberá aceptarse que parte del dinero que gasta la industria farmacéutica en hacer posibles estos ensayos se destina, por ahora, a finalidades tan discutibles como evitables.

### **Competitividad y entorno actual de la investigación farmacéutica en España**

En España, la industria farmacéutica debe recurrir a los centros sanitarios públicos españoles para el desarrollo clínico de nuevos fármacos. Pero, la investigación farmacéutica se internacionaliza cada vez más, y con ella la demanda de ensayos clínicos. Nuestro potencial de investigación clínica cada vez está menos protegido y debe competir con el de otros países de Europa

y del resto del mundo. Conforme en la Unión Europea se generalicen los registros multiestado, dejará de ser preceptiva la realización de muchos de los ensayos que actualmente se realizan en España a requerimiento de nuestras autoridades. La investigación farmacéutica está sometida a la enorme presión de las leyes del mercado y los directores de investigación deben cumplir unos programas y unos calendarios específicos. En estas condiciones, España perderá la poca competitividad que haya podido haber alcanzado si nuestra administración sigue poniendo en juego el calendario de la investigación por demora de los trámites burocráticos, si nuestras instituciones –además de otro tanto de lo mismo– siguen sin estar administrativamente preparadas para contratar e incentivar la realización de ensayos clínicos y, por último, si los investigadores siguen considerando a los ensayos clínicos como una investigación de segunda categoría y no están motivados para ofrecer los mínimos de calidad que se exigen. De seguir esta tendencia, sólo se acudirá a la sanidad española para investigación de trámite y de relevancia dudosa o seudoinvestigación preparada con fines promocionales.

### **El Real Decreto de Ensayos Clínicos de 1993, un paso hacia la normalización de las relaciones económicas**

Entendemos por normalización el que, por una parte, los centros sanitarios capaces de realizar buenos ensayos clínicos sean los más favorecidos y reconocidos y, por otra, que las compañías promotoras compitan para realizar buenos ensayos clínicos en el sector asistencial sanitario español, pagando lo que realmente corresponda por el servicio de investigación que reciben, lógicamente dentro del ámbito de la libre competencia. Con la promulgación de la Ley del Medicamento de 1990<sup>31</sup> y del Real Decreto de Ensayos Clínicos de 1993<sup>32</sup> (en adelante Real Decreto), parecen haberse sentado las bases para alcanzar estos objetivos.

En efecto, tal y como ya ha sido señalado recientemente en esta Revista<sup>33</sup>, mediante el artículo 20 del Real Decreto se pretenden hacer más transparentes las relaciones económicas entre los promotores de un ensayo y los centros sanitarios. Según reza el mencionado artículo: “todos los aspectos económicos relacionados con el ensayo clínico quedarán reflejados en un contrato entre cada centro donde se vaya a realizar el ensayo y el promotor”, además, “en el contrato constará el presupuesto inicial del ensayo es-



pecificando los costes indirectos que aplicará el centro, así como los costes directos extraordinarios, considerando como tales aquellos gastos ajenos a los que hubiera habido si el sujeto no hubiese participado en el ensayo, como análisis y exploraciones complementarias añadidas, cambios en la duración de la atención a los enfermos, reembolso por gastos a los pacientes, compra de aparatos y compensación para los sujetos del ensayo e investigadores. También constarán los términos y plazos de los pagos, así como cualquier otra responsabilidad subsidiaria que contraigan ambas partes”.

En definitiva, esta norma no sólo reconoce la relevancia económica de los ensayos clínicos sino que admite abiertamente la existencia de incentivos para los investigadores. Mediante el presupuesto del ensayo, la institución podrá recuperar los gastos directos extraordinarios originados por la realización del ensayo clínico, y dispondrá de una información tangible para ejercer una política de investigación en el centro. A nadie se le escapa que esta norma va por delante de la realidad por lo que su cumplimiento no estará exento de dificultades. En efecto, las instituciones con tesorería no descentralizada a los centros, deberán modificar su organización administrativa si no quieren aumentar más las incertidumbres de sus investigadores. Incluso en los supuestos que se utilicen figuras administrativas paralelas, todos los centros deberán disponer de un escándalo y saber cuantificar el coste real de los servicios extraordinarios que están en condiciones de prestar. Por otra parte, es de desear que las instituciones no aprovechen la ocasión para establecer gravosos porcentajes en concepto de costes indirectos, encareciendo injustificadamente los ensayos clínicos.

Según el Real Decreto, el contrato económico debe establecerse entre la institución sanitaria y la entidad promotora, sin embargo, la norma no otorga a las instituciones unas responsabilidades específicas para la negociación contractual. Para muchos autores, la existencia de una tercera instancia independiente tanto de los investigadores como del promotor es fundamental<sup>34</sup>. Esta tercera instancia es la que debe garantizar la inexistencia de relaciones al margen del contrato establecido y, sobre todo, es la que debe defender los intereses de los investigadores en planos no estrictamente económicos. Por ejemplo, a través de los contratos también se debería proteger la propiedad intelectual de los investigadores, el derecho de publicación de los resultados y el uso del nombre de los participantes para actividades o fines promocionales.

## **Sobre el destino de los beneficios económicos**

En condiciones normales, los remanentes derivados de un ensayo clínico (ingresos menos gastos reales) deberían tener como principal destino la promoción de nuevas actividades de formación o investigación. Los principales beneficiados deberían ser los propios investigadores involucrados en la realización de ensayos clínicos. La forma y el quién debe gestionar directamente estos fondos requiere una discusión más extensa. Para algunos, en este eslabón también deben establecerse fórmulas para evitar conflictos de intereses, desincentivando el desarrollo de grupos profesionalizados en la realización de ensayos por meros intereses económicos. Se han hecho propuestas de diversos tipos, una de ellas defiende el destino de los remanentes a un fondo común que, controlado por el decano o director de la institución, se destina luego a nuevas actividades de formación o investigación tras concurso o competencia libre entre los distintos investigadores de la institución<sup>19,20</sup>. En cualquier caso, en muchas instituciones sanitarias españolas existen buenos ejemplos de cómo ordenar estos temas de forma razonable.

## **Función de los Comités Éticos de Investigación Clínica**

Entre las funciones que no serán fáciles de ejercer por los Comités Éticos de Investigación Clínica (CEIC) destaca la obligación de estar al corriente de las relaciones económicas derivadas de la realización de los ensayos clínicos. El vacío que existe en España de comités para la integridad de la investigación hará de los CEIC un primer foro de cultura y sensibilización sobre esta problemática, si bien se trata de funciones y responsabilidades que no deben mezclarse.

En el Real Decreto de 1993 se establecen normas específicas tanto para los voluntarios participantes, como para el equipo de investigación y para la propia institución. En cuanto a las relaciones económicas con los voluntarios participantes, el apartado tercero del artículo 20 del Real Decreto dice que “los CEIC podrán conocer el presupuesto del ensayo y evaluarán algunos de los contenidos en los presupuestos tales como las compensaciones para los sujetos participantes en el ensayo”. Es decir, los CEIC deben velar que las compensaciones ofrecidas a los voluntarios sanos no constituyan el atractivo principal para la participación en el ensayo y, por extensión, que los participantes no se profesionalicen como si de alquilar su cuerpo se tratara.

En cuanto a las relaciones económicas con el investigador y su equipo, el CEIC debe conocer y evaluar tanto el alcance de las compensaciones que se ofrecerán a los investigadores como la idoneidad del equipo investigador para el ensayo propuesto, teniendo en cuenta –entre otras cosas– sus obligaciones asistenciales y los compromisos previamente adquiridos con otros protocolos de investigación (apartados quinto y segundo del artículo 42 del Real Decreto). Es decir, si bien la normativa admite la existencia de incentivos para el investigador, parece responsabilizar al CEIC de que los ingresos económicos derivados del ensayo, ni constituyan la principal motivación para realizar el estudio ni fomenten la profesionalización en estas actividades.

Además, la transparencia de las relaciones económicas se refuerza mediante la capacidad que se otorga a los CEIC para requerir información completa sobre las fuentes y cuantía de la financiación del ensayo y sobre toda la distribución de los gastos (artículo 67 del título tercero de la Ley del Medicamento). Esta potestad de los CEIC unida a la presencia del director médico o persona representada por él en su composición, ha de garantizar que el acto de autorización institucional suponga el pleno conocimiento y aceptación de las implicaciones económicas derivadas de la realización de cada ensayo clínico en el centro.

Con el fin de evitar conflictos de intereses entre los miembros de los CEIC, también se han establecido determinadas restricciones e incompatibilidades. En primer lugar, en el apartado cuarto del artículo 4 de la Ley del Medicamento se establece que la pertenencia a los CEIC será incompatible con cualquier clase de intereses derivados de la fabricación y venta de los medicamentos y productos sanitarios. La Orden de Acreditación de los CEIC de la Generalitat de Catalunya<sup>28</sup> obliga a que todo candidato a formar parte de un CEIC declare previamente por escrito a la Administración acerca de todas sus relaciones con la industria sanitaria.

Por último, en el apartado cuarto del artículo 41 del Real Decreto se especifica que ni el CEIC ni ninguno de sus miembros podrán percibir directa ni indirectamente remuneración alguna por parte del promotor del ensayo. Esta restricción se dirige a evitar que los miembros del CEIC no pierdan la necesaria independencia en el ejercicio de sus funciones. El CEIC se debe a la institución del que forma parte y será un asunto interno de la institución la forma y cuantía con que debe incentivar o compensar a sus miembros para que realicen las tareas de la mejor y más eficaz forma posible. Para algunos miembros del CEIC, parti-

cipar con responsabilidad supone un enorme esfuerzo y dedicación adicional a las tareas que ya se tienen encomendadas.

Por ello se comprende la obligación que impone el Real Decreto de que los centros garanticen los medios administrativos y técnicos adecuados así como un presupuesto de funcionamiento (apartado tercero del artículo 41 del Real Decreto).

## Conclusiones

Los conflictos de intereses en investigación clínica han existido, existen y es posible que aumenten en el futuro. El principal problema subyacente es que pensamos que somos ajenos a ellos, con más motivo en tanto que somos un colectivo que aún no tiene la tradición de discutir abiertamente este tipo de situaciones. Su análisis requiere un procedimiento distinto al de los dilemas bioéticos, ya que ningún interés secundario debe prevalecer sobre el principal, la salud de los enfermos y la integridad de la investigación. La mejor forma de prevenir estos conflictos es exponer públicamente los intereses existentes y valorar antes de iniciar la investigación hasta qué punto su integridad se verá comprometida por ellos. En España a la falta de tradición se le suma un inconveniente adicional, la falta de sensibilidad y la escasa organización de los aspectos administrativos y de gestión de la investigación clínica en el sistema asistencial sanitario. En cuanto a la investigación promovida por la industria farmacéutica y sanitaria se refiere, el desarrollo del Real Decreto sobre Ensayos Clínicos de 1993, puede sentar las bases de una progresiva normalización del desorden actualmente existente. Como fuera que además se otorga a los CEIC algunas responsabilidades parciales en la prevención de estos conflictos, existen oportunidades para que nuestro sistema se sensibilice y para que se promueva el debate abierto de nuevas costumbres.

## Agradecimiento

El autor desea agradecer la crítica y las sugerencias recibidas en anteriores versiones del manuscrito por parte de Josep M.<sup>a</sup> Antó, Olav Bakke, Fernando García Alonso, Diego Gracia, Xavier Carné y Miquel Porta.

## BIBLIOGRAFÍA

1. Korenman SG. Conflicts of interest and commercialization o researc Acad Med 1993; 68 (Supl 3): 18-22.

2. Sykes R. Innovación in the pharmaceutical industry. *Br Med J* 1994; 309: 422-423.
3. Pharmaceutical funds for clinical research: a mixed blessing. *Lancet* 1987; 1: 257-258.
4. Thompson DF. Understanding financial conflicts of interest. *N Engl J Med* 1993; 329: 573-576.
5. Rothman KJ. Conflict of interest. The new McCarthyism in Science. *JAMA* 1993; 269: 2.782-2.784.
6. Witt MD, Gostin LO. Conflict of interest dilemmas in biomedical research. *JAMA* 1994; 271: 547-551.
7. Relman AS. Economic incentives in clinical investigation. *N Engl J Med* 1989; 320: 933-934.
8. Brennan TA. Bying editorials. *N Engl J Med* 1994; 331: 673-675.
9. PHS Grants Policy Statement, Revised. Washington, DC: US Public Health Service 1991. Dept. of Health and Human Services publication 82-50,000.
10. Resquest for comment on proposed guidelines for policies on conflicts of interest. NIH Guide to Grants and Cochracts. Bethesda, MD: National Institutos of Health 1989; 18 (No. 32): 3-4. (La versión oficial corregida ha sido publicada en el Federal Register de 28 de junio de 1994.)
11. Harvard University, Faculty of Medicine. Policy on conflicts of interest and commitment. Boston: Harvard Medical School, marzo 22, 1990.
12. Council on Scientific Affairs and Council on Ethical and Judicial Affairs. American Medical Association. Conflicts of interest in medical center/ industry research relationships. *JAMA* 1990; 263: 2.790-2.793.
13. American Federation for Clinical Research guideliness for avoiding conflict of interest. *J Clin Res* 1990; 38: 239-240.
14. Guidelines for dealing with faculty conflicts of commitment and conflicts of interest in research. *Acad Med* 1990; 65: 485-495.
15. Gunsalus CK. Institutional structure to ensure research integrity. *Acad Med* 1993; 68 Supl 3: 33-38.
16. Camí J. Promoción de medicamentos, promoción científica y conflicto de intereses. *Gac Sanit. En prensa.*
17. Freestone DS, Mitchell H. Inappropriate publication of trial results and potential for allegations of illegal share dealing. *Br Med J* 1993; 306: 1.112-1.114.
18. Booth W. Conflict of interest eyed at Harvard. *Science* 1988; 242: 1.497-1.499.
19. Shimm DS, Spece RG. Industry reimbursement for entering patients into clinical trials: legall and ethical issues. *Ann Intern Med* 1991; 115: 148-151.
20. Shimm DS, Spece RG. Ethical issues and clinical triais. *Drugs* 1993; 46: 579-584.
21. Internacional Committee of Medical Journal Editors. Conflict of interest. *Lancet* 1993; 341: 742-743.
22. Relman AS. Dealing with conflict of interest. *N Engl J Med* 1984; 310: 1.182-1.183.
23. Kassirer JP, Angell M. Financial conflicts of interest in biomedical research. *N Engl J Med* 1993; 329: 570-571.
24. Kassirer JP, Angell M. The journal's policy on cost-effectiveness analyses. *N Engl J Med* 1994; 331: 669-670.
25. Smith R. Conflict of interest and the *Br Med J*. *Br Med J* 1994; 308: 4-5.
26. Topol EJ, Armstrong P, Van der Werf F, Kleiman N, Lee K, Morris D et al. Confronting the issues of patient safety and investigator conflict of interest in an internacional clinical trial of myocardial reperfusion. *J Am Coll Cardiol* 1992; 19: 1.123-1.128.
27. The GUSTO Investigators. An internacional randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993; 329: 673-682.
28. Friedman MA, McCabe MS. Assigning care costs associated with therapeutic oncology research: a modest proposal. *J Natl Cancer Inst* 1992; 84: 760-763.
29. Stamler J. Effective clinical triais investigators and adequate incentivos. *Clin Pharmacol Ther* 1979; 25: 666-670.
30. Buchwald H. NHLBI clinical trials and conflicts of interest. *Controlled Clinical Trials* 1990; 11: 217-222.
31. Ley 25/1990 de 20 de diciembre, del Medicamento. BOE de 22 de diciembre de 1990; 306: 38.228-38.246.
32. Ministerio de Sanidad y Consumo. Real Decreto 561/1993, de 16 de abril, por el que se establecen los requisitos para la realización de ensayos clínicos con medicamentos. BOE de 13 de mayo de 1993; 114: 14.346-14.364.
33. García López FJ, Pijoán Zubizarreta JI. Comentarios a la nueva regulación legal de los ensayos clínicos. *Med Clin (Barc)* 1994; 103: 587-590.
34. Chren MM. Independent investigators and for-profit companies. Guidelines for biomedical scientists considering funding by industry. *Arch Dermatol* 1994; 130: 432-437.
35. Departament de Sanitat i Seguretat Social. Ordre de 26 d'octubre de 1992, d'acreditació dels comitès ètics d'investigació clínica. DOG de 18 de noviembre de 1992; 1.671: 6.679-6.680.

## Parte 5

---

# **CONSENTIMIENTO INFORMADO**

---

---

# Investigación clínica con fármacos: estudio de la influencia de la información sobre reacciones adversas en la obtención del consentimiento informado

---

Rafael Dal-Ré

Departamento de Investigación Clínica. Smith Kline Beecham Pharmaceuticals. Madrid.

A lo largo de las últimas décadas se han realizado numerosos estudios empíricos con objeto de conocer las percepciones que los individuos participantes en ensayos clínicos (EC) tienen sobre diversos aspectos del proceso del consentimiento informado. La gran mayoría de ellos se han realizado en EE.UU. donde es preceptivo utilizar el documento que recoja la información mínima a leer y entregar al paciente<sup>1</sup>. La cantidad de información aportada debe ser aquella con la que se considere que una persona razonable pueda tomar una decisión juiciosa<sup>1,2</sup>.

Uno de los elementos de información a aportar a los sujetos de un EC según la normativa española vigente<sup>3</sup> es el de los riesgos predecibles que aquél conlleva. Desde 1989 la Dirección General de Farmacia y Productos Sanitarios solicita la redacción de un documento de información mínima para ser utilizado en todos los enfermos incluidos en un EC aprobado por ella.

Si bien se ha sugerido<sup>4</sup> que aportar información detallada acerca de los riesgos puede conducir a muchos enfermos a rechazar su participación en el EC propuesto, lo cierto es que no se ha realizado ningún estudio encaminado a evaluar específicamente este extremo. El objetivo de este estudio es evaluar la importancia que la cantidad de información referente a reacciones adversas tiene en el rechazo o aceptación por los individuos a los que se propone participar en un EC.

## Sujetos y método

---

Este estudio está basado en una situación experimental simulada. A los individuos (estudiantes universitarios) se les pidió su opinión acerca de si querían o no participar en un EC que, sin ellos saberlo, era ficticio. Para que el EC (ficticio) propuesto fuese coherente con el objetivo del

estudio, se consideró que debía tener las características siguientes: *a)* ser adecuado para realizarlo en voluntarios sanos, es decir, en fase 1; *b)* no producir molestias (procedimientos invasivos) a los individuos, para evitar que el posible rechazo se debiese a esta causa; *c)* requerir del voluntario un tiempo limitado; *d)* "realizarlo" con un fármaco en investigación, es decir, no comercializado; *e)* abonar una cantidad de dinero, habitual en estudios en fase 1, que no fuese *per se* una razón para aceptar participar. Por todo ello, se planteó un "EC" cuyo fin fuese estudiar el metabolismo de un fármaco a través de la toma de una muestra urinaria tras la administración oral de una sola dosis del fármaco. El fármaco al que se denominó DRS-8189 debería poseer ciertas características que lo hicieran susceptible de un ensayo de este tipo. El DRS-8189 no es otro que la isoniacida. A los individuos participantes siempre se les habló del DRS-8189 como un fármaco en investigación, no comercializado en España. Si bien parte de la información que se aportó a los estudiantes era falsa, la referida a las diferencias en su metabolismo según las razas, a su interacción con el alcohol y, esto es la esencia del estudio, a las reacciones adversas, era la de la isoniacida<sup>5</sup>. En el "EC", al estar el fármaco en fase de investigación, no se podía incluir ninguna mujer embarazada o en período de lactancia y así se comunicó a los sujetos. Por último, se informó a los posibles participantes que recibirían 1.000 ptas. como gratificación, cantidad que podían emplear en trasladarse en taxi al lugar donde el estudio se llevaría a cabo, el Hospital Clínico de Madrid. De esta forma se cumplían las cinco condiciones mencionadas anteriormente.

Para conseguir el objetivo del estudio, se redactaron en lenguaje llano los documentos de información mínima (documentos A y B) que constaban de cinco apartados. Cuatro de ellos (Información previa, Advertencia, Procedimiento y Realización del estudio) eran literalmente idé-

TABLA I  
 CARACTERÍSTICAS DEMOGRÁFICAS DE LOS GRUPOS DE INDIVIDUOS  
 A LOS QUE SE HA PROPUESTO PARTICIPAR EN EL ESTUDIO

	Grupo A <sup>a</sup>			Grupo B <sup>b</sup>		
	Total	Facultad de medicina	Teleco	Total	Facultad de medicina	Teleco
N.º de individuos	160	79	81	146	69	77
Sexo (V/M)	91/69	29/50	62/19	80/66	28/41	52/25
Edad (años)						
$\bar{x} \pm DE$	19,9 $\pm$ 1,4	20,5 $\pm$ 1,5	19,3 $\pm$ 0,7	19,9 $\pm$ 1,3	20,4 $\pm$ 1,3	19,4 $\pm$ 0,9
Intervalo	18-32	19-32	18-22	18-30	19-30	18-24

<sup>a</sup> Se les aportó la información del documento A (Vademécum<sup>6</sup>).

<sup>b</sup> Se les aportó la información del documento B (PDR<sup>7</sup>).

Teleco = Escuela Técnica Superior de Ingenieros de Telecomunicaciones; V = varón; M = mujer.

tics en los dos documentos. El restante, referente a los efectos secundarios era, lógicamente diferente. En efecto, en el documento A se recogieron como reacciones adversas del DRS-8189 las incluidas en el Vademécum español<sup>6</sup> para la isoniácida; en el documento B, se incluyeron la gran mayoría de las reacciones adversas recogidas en el Vademécum norteamericano (Physicians Desk Reference, PDR)<sup>7</sup>. En ambos, se incluye una información sobre el porcentaje global de efectos secundarios asociados a isoniácida y, en el documento B, los porcentajes de hepatotoxicidad según la edad de los pacientes, recogidos en un texto clásico de farmacología<sup>5</sup>.

Para este estudio se eligieron estudiantes pertenecientes a la Facultad de Medicina de la Universidad Complutense, y de la Escuela Técnica Superior de Ingenieros de Telecomunicaciones de la Universidad Politécnica de Madrid.

Aprovechando la presencia de los alumnos en clase, en cada centro universitario se celebraron dos reuniones en noviembre de 1989, que duraron unos 20 minutos cada una de ellas. A una brevísima presentación sobre la necesidad de realizar EC en el desarrollo de todo fármaco, siguió la lectura del documento (A o B) asignado a cada uno de los grupos (A o B); a su conclusión se contestaron algunas preguntas formuladas por parte de los estudiantes y se repartió un documento a cada alumno instándoles a que lo leyeran con detenimiento. A cada estudiante se le pidió que de forma anónima contestase en el propio documento si quería o no participar, y que indicase la edad y el sexo. A todos ellos se les dijo que en unas semanas se darían los detalles para que aquellos que quisieran participar en el "EC" se personasen en el Hospital Clínico en una fecha y hora determinadas.

En la Facultad de Medicina se realizó la encuesta antes de que se impartiesen las clases

referentes a los tuberculostáticos, con el fin de evitar que identificasen al DRS-8189 con la isoniácida. En ambos centros académicos, se leyó y entregó el documento A en el primer día del estudio, y el documento B en el segundo.

Transcurridas 2 semanas, se informó por escrito a los profesores que habían colaborado en el estudio, con objeto de que comunicasen a sus alumnos, que, lamentablemente, el "EC" no iba a realizarse debido al reducido número de voluntarios.

El estudio fue aprobado por el Comité de Ensayos Clínicos del Hospital Clínico de Madrid, información incluida en los dos documentos utilizados.

En el análisis estadístico se empleó de la ji al cuadrado ( $\chi^2$ ) y, cuando ésta lo requirió, la prueba exacta de Fisher.

## Resultados

Un total de 306 estudiantes participaron en este estudio. El 48 % eran estudiantes de Medicina y el 56 % eran varones. La distribución por sexos mostró notables diferencias entre los universitarios de uno y otro centro académico: el 61,5 % de los estudiantes de Medicina eran mujeres, mientras que en Telecomunicaciones el 72 % eran varones. La edad media ( $\pm$  DE) fue de 19,9 ( $\pm$  1,3) años.

En la tabla I se muestran las características demográficas de los grupos, A y B, que recibieron la información del documento A y del documento B, respectivamente. En ambos grupos el tanto por ciento de estudiantes procedentes de cada uno de los dos centros universitarios fue similar. Del mismo modo, también resultó similar la distribución de sexos y de la edad media de ambos grupos.

El 10,6 % de los individuos del grupo A aceptaron participar en el "EC", frente al 4,1 % del grupo B, diferencia que fue estadísticamente significativa (tabla II). Al estudiar la distribución de voluntarios según el grupo y el centro universitario, no se encontraron diferencias estadísticamente significativas (tabla III). Cabe resaltar que entre las mujeres, sólo entre las estudiantes de Medicina hubo quienes estaban dispuestas en intervenir en el "EC"; además, el 12 % del grupo B participaría en el "EC", frente al 8 % de los del grupo A.

Por último, la distribución de los individuos que aceptaron y rechazaron participar en el "EC" según el centro universitario mostró que entre los estudiantes de Medicina hubo una proporción significativamente mayor de voluntarios que entre los de Telecomunicaciones: 17 de 148 (11,5 %) frente a 6 de 158 (3,8 %), respectivamente ( $\chi^2 = 6,49, p < 0,05$ ).

**Discusión**

En diversas ocasiones se han empleado situaciones similares para evaluar diferentes aspectos relacionados con el consentimiento del enfermo en los EC. Así, McNeil et al<sup>8</sup> pidieron a más de mil individuos (enfermos, médicos y graduados universitarios) que supusieran que padecían cáncer de pulmón, para evaluar en qué medida la información que se aporta a los enfermos influye en su decisión acerca de terapias alternativas. Bigorra y Baños<sup>9</sup>, a través de un cuestionario, preguntaron a 250 estudiantes de Medicina si estarían dispuestos a participar en un ensayo en fase 1 y cuáles eran las razones para ello. Un "ensayo" similar al de este estudio, en el sentido de que los participantes ignoraban que se les pedía participar en un EC ficticio, lo realizaron Epstein y Lasagna<sup>4</sup>. Estos autores solicitaron a 66 trabajadores hospitalarios su participación en un "EC" doble ciego, para valorar la eficacia analgésica en el tratamiento de la cefalea de un fármaco, de nombre ficticio, frente a placebo. Para realizar este estudio, redactaron tres documentos de información para el paciente que diferían entre sí en la cantidad de información sobre el "fármaco" (la aspirina), especialmente aquella referente a la descripción de las acciones, efectos secundarios y toxicidad. Epstein y Lasagna<sup>4</sup> mostraron, entre otras cosas, que la aceptación de consentir participar se encuentra "inversamente relacionada con la longitud del documento". Los resultados de este estudio muestran que a mayor cantidad de información referente a las reacciones adversas asociadas al empleo de un fármaco, menor acep-

TABLA II  
ACEPTACIÓN Y RECHAZO A PARTICIPAR EN EL " ENSAYO CLÍNICO " EN LOS DOS GRUPOS DE INDIVIDUOS

	<i>Si</i>	<i>No</i>	<i>Significación estadística (p)</i>
Grupo A <sup>a</sup>	17	143	
Grupo B <sup>b</sup>	6	140	< 0,05

<sup>a</sup> Se les aportó la información del documento A (Vademécum<sup>6</sup>).

<sup>b</sup> Se les aportó la información del documento B (PDR<sup>7</sup>).

TABLA III  
DISTRIBUCIÓN DE LOS INDIVIDUOS DE LOS DOS GRUPOS QUE ACEPTARON PARTICIPAR EN EL " ENSAYO CLÍNICO " SEGÚN GRUPO, CENTRO UNIVERSITARIO Y SEXO

	<i>Grupo A<sup>a</sup></i>	<i>Grupo B<sup>b</sup></i>
Facultad de Medicina		
N.º de individuos	12	5
Sexo (V/M)	8/4	0/5
Teleco		
N.º de individuos	5	1
Sexo (V/M)	5/0	1/0

<sup>a</sup> Se les aportó la información del documento A (Vademécum<sup>6</sup>).

<sup>b</sup> Se les aportó la información del documento B (PDR<sup>7</sup>).

V = varón; M = mujer; Teleco = Escuela Técnica Superior de Ingenieros en Telecomunicaciones.

tación en consentir entre los posibles participantes: concretamente, disminuyó en el 61 %. Aunque los resultados coinciden con los de Epstein y Lasagna<sup>4</sup>, hay que observarlos con cautela. En efecto, este estudio tiene algunas particularidades que le alejan claramente de la práctica investigadora en pacientes. En primer lugar, los individuos del estudio eran sanos y, prácticamente, de la misma edad. Además, todos eran universitarios. El procedimiento utilizado en la lectura y posterior entrega del documento de información mínima no tiene nada que ver con la forma habitual de proceder en los EC; en éstos, el investigador suele estar a solas con el paciente en presencia, quizá, de un testigo y/o algún colaborador del EC. Sin embargo, en este estudio se utilizó el sistema recomendado (lectura y entrega de información) por el 76 % de los más de cien miembros de Comités de EC encuestados en

nuestro país<sup>10</sup> y la decisión de participar o no la tuvo que hacer el individuo en la misma sesión en la que se le informó la forma más común de proceder en la práctica<sup>11</sup>.

Los resultados en este estudio deberían ser confirmados en EC con pacientes; sin embargo, esto es una tarea casi imposible de realizar y, quizá por ello, nadie, desde que se publicó el estudio de Epstein y Lasagna<sup>4</sup> en 1969, la ha abordado todavía. Piénsese en la gran cantidad de tiempo empleado por parte de los enfermos y, sobre todo, por el personal investigador, para que al final, se tenga que informar a los pacientes que el EC no se va a realizar, que era ficticio. Además, los enfermos no podrían estar esperando durante un tiempo indeterminado a recibir el "tratamiento" en investigación. Otra alternativa para investigar la influencia de información apodada con respecto a las reacciones adversas, en la voluntariedad de los pacientes en consentir participar en un EC, sería hacerlo en el marco de un ensayo real. Ahora bien, si sabido es que la tasa de reclutamiento de enfermos de los EC es siempre menor a la esperada<sup>12</sup> ¿qué investigador se atrevería a utilizar, por ejemplo, dos documentos de información mínima que varíen sustancialmente de la información referente a las reacciones adversas? Los resultados de este estudio y los de Epstein y Lasagna<sup>4</sup> sugieren que el rechazo a participar sería mayor entre los enfermos que recibiesen más cantidad de información, lo que podrían conducir a que el ensayo se prolongase excesivamente. Informar al enfermo potencial acerca de los riesgos del EC y, claro está, de las reacciones adversas previsible, está recogido en la Declaración de Helsinki, a pesar de los cual no siempre se hace<sup>13,14</sup>.

Además, en nuestro país el 92 % de los miembros de Comités de EC encuestados<sup>10</sup> considera que siempre se debe informar sobre los riesgos previsible a los enfermos participantes en EC. Si en la actualidad el investigador español debe utilizar un documento que recoja la información mínima a aportar a los enfermos de un EC aprobado por la Administración y un Comité de EC, ¿qué cantidad de información debe incluir en ese documento? Hay quienes piensan que es inadecuado informar de todos los riesgos si ello puede conducir al enfermo a no participar en un EC que le puede ser beneficioso<sup>15</sup>. Asimismo, parece innecesario detallar los efectos secundarios que se presentan con una muy baja incidencia<sup>16</sup>. Suponiendo que el EC objeto de este estudio fuese real y que un investigador tuviese que llevarlo a cabo, ¿qué información referente a las reacciones adversas sería más correcto que utilizase, la del do-

cumento A o la del B? ¿Lo más idóneo sería encontrar un equilibrio entre ambos? Para intentar evitar, en alguna medida, que un investigador incluya en los documentos de información mínima a aportar a los pacientes efectos secundarios que otro omite, sería conveniente que se emitiesen unas recomendaciones generales que orientasen a investigadores y Comités de EC a este respecto. Un Comité Ético Nacional, cuya creación ya ha sido propuesta<sup>17</sup>, quizá podría realizar esta labor.

## Agradecimientos

El autor agradece a los Prof. P. Lorenzo y J. Tamargo de la Facultad de Medicina (UCM), al Prof. I. Esquivias de la E.T.S. de Ingenieros de Telecomunicaciones (UPM) y al Prof. A. Moreno del Hospital Clínico de San Carlos (Madrid), la colaboración prestada para la realización de este estudio.

## BIBLIOGRAFÍA

1. Levine RJ. Ethics and Regulation of Clinical Research, 2nd ed. Baltimore: Urban and Schwarzenberg, 1986.
2. Gracia D. Fundamentos de Bioética. Madrid: EUEDEMA SA, 1989.
3. Orden. Ministerio de Sanidad y Consumo, 3 de agosto de 1982. Marid: BOE, 12 de agosto de 1982.
4. Epstein LC, Lasagna L. Obtaining informed consent. Arch intern Med 1969; 123: 682-687.
5. Mandel GL, Sande MA. Drogas usadas en la quimioterapia de la tuberculosis y de la lepra. En: Gilman AG, Goodman LS, Rall TW, Murad F, eds. Las Bases Farmacológicas de la Terapéutica, 7.ª ed. México: Editorial Médica Panamericana SA, 1986; 1.142-1.159.
6. Vademécum Internacional, 30.ª ed. Madrid: Medicom SA, 1989.
7. Physicians' Desk Reference, 43th ed. Oradell, NJ: Medical Economics Company Inc, 1989.
8. McNeil BJ, Pankar SQ, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med 1982; 306: 1.259-1.262.
9. Bigorra J, Baños JE. Weight of financial reward in the decision by medical students and experienced healthy volunteers to participate in clinical trials. Eur J Clin Pharmacol 1990; 38: 443-446.
10. Dal-Ré R. Informed consent in clinical research with grugs in Spain: perspective of Clinical Trials Committees members. Eur J Clin Pharmacol 1990; 38: 319-324.
11. Lewis JM. An IRB's consent form survey. IRB 1984; 6: 10.
12. Friedman LM, Furberg CO, De Mets DL. Fundamentals of clinical trials, 2nd ed. Littleton: PSG Publishing Co Inc, 1985.



13. Gray BH, Cooke RA, Tannenbaum AS. Research involving human subjects. *Science* 1978; 201: 1.094-1.101.
14. Riecken HW, Ravish R. Informed consent to biomedical research in Veterans Administration Hospitals. *JAMA* 1982; 248: 344-348.
15. Brahams D. Informed consent does not demand full disclosure risks. *Lancet* 1983; 2: 58.
16. Dukes MN, Swartz B. Responsibility for drug-induced injury. Amsterdam: Elsevier Science Publishers B.V., 1988.
17. Serrano MA, García Alonso F, González de Suso MJ et al. Ensayos clínicos en España (1982-1988). Serie Monografías Técnicas de la Dirección General de Farmacia y Productos Sanitarios, nº 17. Madrid: Ed. Ministerio de Sanidad y Consumo, 1990.

## Apéndice

### (DOCUMENTO A)

#### INFORMACIÓN SOBRE EL ESTUDIO DEL METABOLISMO DEL DRS-8189 EN VOLUNTARIOS SANOS DE ETNIAS MEDITERRÁNEAS

##### 1. INFORMACIÓN PREVIA

El DRS-8189 es un agente antiinfeccioso investigado y desarrollado en Japón. Este fármaco no está comercializado en España. Los estudios realizados en voluntarios sanos y enfermos japoneses y escandinavos han demostrado la eficacia y seguridad del DRS-8189. Sin embargo, se ha visto que, como ocurre con otros fármacos de estructura química similar, las distintas poblaciones humanas muestran diferencias genéticas en cuanto al metabolismo del DRS-8189. Se ha comprobado que la raza oriental (japonesa) presenta diferencias sustanciales con respecto a los escandinavos; es de esperar que también existan diferencias entre éstos y la raza mediterránea. El estudio para el que pedimos su participación, tiene por objeto conocer cuál es el metabolismo del DRS-8189 en la raza mediterránea, dato de importancia capital para decidir cuál será la dosis más idónea para los futuros enfermos españoles.

##### 2. ADVERTENCIA

Cualquier persona sana puede participar en este estudio. No lo podrán hacer las mujeres embarazadas o en período de lactancia.

##### 3. EFECTOS SECUNDARIOS

La frecuencia de los efectos secundarios a la administración repetida de DRS-8189 en más de 2.000 pacientes tratados es del 5,4%. La dosis habitual recibida por estos enfermos fue de hasta 300 mg/día. Los efectos secundarios son benignos y pasajeros: dolor de cabeza, vértigo, estreñimiento o diarreas, erupciones de la piel fugaces.

##### 4. PROCEDIMIENTO

Si Ud. decide participar, deberá acudir en el día y hora previamente señalados al Servicio de Farmacología Clínica del Hospital Clínico de Madrid. Deberá leer nuevamente esta información y podrá formular cualquier pregunta a los investigadores del estudio. Veinticuatro horas después, y siguiendo la legislación española vigente (BOE, 12 de agosto 1982), deberá firmar junto con los investigadores del estudio, un documento en el que expresará que su participación es voluntaria. Se le darán las instrucciones pertinentes por escrito. En esencia éstas son: a) se le entregará un comprimido de 50 mg de DRS-8189, que deberá tomar durante la cena del día acordado; b) deberá recoger 20 ml de la primera orina de la mañana siguiente a la toma del fármaco, en un envase estéril que se le proveerá a tal efecto; c) deberá abstenerse de tomar alcohol durante las 24 h previas y las 24 h posteriores a la toma del fármaco. A cada participante en este estudio se le gratificará con 1.000 ptas.

##### 5. REALIZACIÓN DEL ESTUDIO

Correrá a cargo del Prof. A. Moreno González, Servicio de Farmacología Clínica, Hospital Clínico S. Carlos, 5ª planta, Ala Sur, Madrid. Este estudio ha sido aprobado por el Comité de Ensayos Clínicos del Hospital Clínico San Carlos. Si Ud. decide participar en este estudio, tenga en cuenta que su participación es voluntaria y que en cualquier momento, incluso después de firmar el documento al que se ha hecho referencia anteriormente, Ud. puede abandonar el estudio sin tener que dar explicación alguna.

(Señale lo que proceda)

Sexo: Varón [  ] Mujer [  ] Edad (años): \_\_\_\_\_

¿Desearía participar en este estudio?: Sí [  ] No [  ]

**(DOCUMENTO B)**

Contenido idéntico al del documento A salvo el apartado 3.

**3. EFECTOS SECUNDARIOS**

La frecuencia de efectos secundarios a la administración repetida de DRS-8189 en más de 2.000 pacientes tratados es del 5,4 %. La dosis habitual recibida por estos enfermos fue de hasta 300 mg/día.

Los efectos secundarios al DRS-8189 son:

*Sistema nervioso:* sensación de hormigueo en manos y pies; en raras ocasiones, deterioro transitorio de la memoria y reacciones psicóticas.

*Aparato digestivo:* náuseas, vómitos, molestias estomacales.

*Hígado:* hepatitis tras la administración del fármaco durante varias semanas. En la mayor parte de las ocasiones la hepatitis es reversible. El daño hepático es raro en individuos menores de 20 años; del 0,3 % en pacientes entre 20 y 34 años; del 1,2 % en enfermos de 35 a 49 años y del 2,3 % en mayores de 50 años.

*Sangre:* disminución de los glóbulos rojos (anemia) y plaquetas (trombocitopenia).

*Reacciones alérgicas:* fiebre, erupciones en la piel, inflamación de los vasos sanguíneos.

*Reacciones endocrinas y metabólicas:* ginecomastia (desarrollo excesivo de la mama en el varón), deficiencia de vitamina B y aumento de la concentración de glucosa en sangre.

# El consentimiento informado: teoría y práctica (I)

Pablo Simón Lorda<sup>a</sup> y Luis Concheiro Carro<sup>b</sup>

<sup>a</sup>Médico de Medicina Familiar y Comunitaria. Hospital Ramón y Cajal. Alumno del Máster de Bioética de la Universidad Complutense. <sup>b</sup>Catedrático de Medicina Legal. Facultad de Medicina. Universidad de Santiago de Compostela. Santiago de Compostela. La Coruña.

A comienzos de 1991 el Centro de Investigaciones Sociológicas (CIS) hizo públicos los resultados de una encuesta realizada en enero de 1990 para valorar el grado de satisfacción e insatisfacción de los usuarios de los servicios sanitarios públicos de nuestro país<sup>1</sup>. Entre los muchos datos que proporciona la encuesta merece la pena destacar los referidos a la información al usuario en los ámbitos de la atención primaria, asistencia hospitalaria y servicios de urgencias.

Preguntados los usuarios acerca de cómo creen ellos que se entiende la información que proporcionan los médicos de asistencia primaria, un 43 % estima que se entiende "regular o mal", frente a un 45 % que cree que "bien o muy bien".

En el ámbito hospitalario las cifras son del 41 % que cree que la información se entiende "regular, mal o muy mal", frente al 42 % que opina que "bien o muy bien".

En el caso de los servicios de urgencia, un 31 % de los usuarios estima que la información proporcionada por los médicos de urgencias es "regular, mala o muy mala", frente a un 44 % que la califica como "buena o muy buena".

Estos datos sugieren, pues, que el flujo de información desde los médicos hacia los usuarios no es, en conjunto, satisfactorio. Las necesidades de información de los pacientes son por tanto mucho mayores de lo que habitualmente estiman los médicos. Estas necesidades no son un capricho de los enfermos sino que forman parte de un derecho que ha sido recogido en el artículo 10 de la Ley General de Sanidad de 25 de abril de 1986: el derecho al consentimiento informado. El artículo en cuestión recoge la *Carta de Derechos y Deberes de los Enfermos*, y es en los puntos 5, 6 y 9 donde se describe este derecho.

Lo cierto es que a pesar de estar vigente desde 1986 esta carta no ha sido suficientemente di-

fundida y, por lo tanto, es todavía muy poco conocida por los usuarios. De hecho en la encuesta antes citada, ante la pregunta "¿ha oído hablar usted de la Carta de Derechos y Deberes de los Enfermos?", nada más y nada menos que un 78 % responde que NO, frente a tan sólo un 21 % que dice que SÍ.

Resulta por lo tanto imprescindible y urgente un cambio de actitud de los profesionales sanitarios hacia los derechos de los pacientes, y en particular hacia su derecho al consentimiento informado. Un primer paso para avanzar en esta dirección podría ser conocer mejor en qué consiste este derecho y sugerir estrategias para respetarlo de forma más satisfactoria.

## El derecho al consentimiento informado

Una primera afirmación importante a realizar es que este derecho no es una ficción legal, sino que ha sido considerado como un derecho *humano*. Hablamos pues de un imperativo ético, que además en nuestro país, como en muchos otros, ha cristalizado como exigencia legal. Las exigencias legales representan a su vez *mínimos éticos* —que por desgracia muchas veces caen "bajo *mínimos*", tal y como en parte le ha sucedido a la forma en que está redactada la Carta de Derechos de la ley española—. Sin embargo, un buen profesional no debería contentarse sin más con mínimos, sino que debería además tratar de elevar mucho más los horizontes éticos de sus actuaciones, esto es, de *maximizarlos*.

La teoría del consentimiento informado es de origen anglosajón, más concretamente norteamericano, aunque después ha pasado a formar parte del patrimonio ético de toda la práctica médica occidental. Si bien su gestación tiene mucho que ver con el propio modelo ético-político con que aquel país vio la luz a finales del siglo XVIII, su verdadero desarrollo hay que situarlo en el contexto del amplio movimiento de reivindicación de los de-

rechos civiles que, iniciándose a finales de la Segunda Guerra Mundial, tiene su auge en la década de los sesenta-setenta –y que tampoco se limita sólo a EE.UU.–. Este movimiento, junto con otros factores como por ejemplo el acelerado desarrollo científicotécnico de la medicina, repercute en el ámbito sanitario de tres maneras que están a su vez íntimamente relacionadas. En primer lugar impulsa la reivindicación de las cartas de derechos de los enfermos, de los que quizás el más importante es precisamente el derecho al consentimiento informado, cuyo punto de partida más inmediato es el Código de Nüremberg de todos conocido. De hecho una de las áreas de la medicina donde mayor impulso ha tomado la teoría del consentimiento informado es la de la investigación. En segundo lugar potencia el desarrollo de una nueva disciplina que va a tener una marcada influencia en el modo de actuación de los médicos estadounidenses: *la bioética*. A su vez los avances de esta nueva disciplina van a estar estrechamente relacionados con el desarrollo de la teoría del consentimiento informado en el ámbito jurídico. En tercer lugar hace entrar definitivamente en crisis el modelo de relación médico-paciente clásico, que estaba marcado por el paternalismo y que consideraba al paciente no sólo como un minusválido físico, sino también psicológico-moral, y por tanto lo creía incapaz de tomar decisiones por sí mismo. Frente a este modelo clásico irrumpe ahora con fuerza otro que pone su acento en la autonomía de los propios enfermos para decidir lo que debe o no debe hacerse con ellos. El corazón de este nuevo modelo va a ser precisamente la teoría del consentimiento informado.

### Los contenidos mínimos del consentimiento informado

La edición de 1984 del “Manual de Ética” del Colegio de Médicos Americanos definía el consentimiento informado de la siguiente manera<sup>2</sup>:

“El consentimiento informado consiste en la explicación, a un paciente atento y mentalmente competente, de la naturaleza de su enfermedad, así como del balance entre los efectos de la misma y los riesgos y beneficios de los procedimientos terapéuticos recomendados, para a continuación solicitarle su aprobación para ser sometido a esos procedimientos. La presentación de la información al paciente debe ser comprensible y no sesgada; la colaboración del paciente debe ser conseguida sin coerción; el médico no debe sacar partido de su potencial dominancia psicológica sobre el paciente.”

Esta definición tan sólo presenta un fallo que es necesario destacar: que donde pone “procedimientos terapéuticos” debería poner más bien “procedimientos diagnósticos y terapéuticos”. Pero es una definición que consideramos muy importante porque en ella se encuentran presentes casi todos los elementos que componen la teoría del consentimiento informado. Estos son los que a continuación se reseñan.

#### *Voluntariedad*

El consentimiento, informado o no, para un determinado procedimiento diagnóstico o terapéutico, que es emitido por un paciente que no actúa de forma voluntaria, no es aceptable ni desde el punto de vista ético ni desde el legal. Según la President’s Commission, la limitación de la libertad puede presentarse de tres formas posibles<sup>3</sup>:

*Persuasión.* En este caso –quizás el más frecuente de todos– el paciente es sometido a un procedimiento sin darle la oportunidad de que efectúe ningún tipo de elección. Aunque este tipo de actuaciones pueden ser ética y socialmente aceptables en determinadas ocasiones, tales como las situaciones de amenaza de la salud pública, en la práctica común no lo son.

*Coacción.* A diferencia del caso anterior en que no hubo obtención de consentimiento alguno, en este caso el consentimiento del paciente sí se consigue, pero bajo coacción. Una decisión está tomada bajo coacción cuando la persona está amenazada por otra de forma explícita o implícita, con consecuencias no deseadas y evitables si accede a sus requerimientos. La coacción será tanto más potente cuanto mayor diferencia exista entre el poder del coaccionador y el del paciente, puesto que a mayor poder de aquél, mayor credibilidad tendrán sus amenazas para éste. Esta coacción puede ser ejercida por el médico, la enfermera, un familiar con intereses propios, etc. Por ejemplo, la amenaza del “alta voluntaria” que se contempla en el artículo 10 de la Ley General de Sanidad en caso de negativa al tratamiento podría ser una forma de coacción.

*Manipulación.* El médico, por sus conocimientos e influencia psicológica, se encuentra en una posición respecto al paciente que le capacita para presentarle la información de tal manera que le empuje a tomar una decisión determinada. Esta conducta, si es deliberada y basada en la distorsión sesgada y fraudulenta de la informa-

ción, supone una anulación del requisito de voluntariedad.

### *Información en cantidad suficiente*

Para que un “consentimiento” se considere aceptable no sólo debe de ser “libre”, sino también “informado”, lo que quiere decir que tiene que ser emitido tras un proceso de evaluación de una determinada cantidad de información relativa a la decisión a tomar.

El punto 5 del artículo 10 de la Ley General de Sanidad dice al respecto que el paciente tiene derecho “a que se le dé, en términos comprensibles, a él y a sus familiares o allegados, información completa y continuada, verbal y escrita sobre su proceso, incluyendo diagnóstico, pronóstico y alternativas de tratamiento”. De una forma más amplia podríamos decir que los ámbitos que debería abarcar la información proporcionada, tal y como lo entiende la teoría del consentimiento informado estadounidense, son los siguientes<sup>4,5</sup>:

1. Descripción del procedimiento Propuesto, tanto de sus objetivos como de la manera en que se llevará a cabo.

2. Riesgos, molestias y efectos secundarios posibles.

3. Beneficios del procedimiento a corto, medio y largo plazo.

4. Posibles procedimientos alternativos con sus riesgos, molestias y efectos secundarios respectivos, y explicación de los criterios que han guiado al médico en su decisión de recomendar el elegido en lugar de estos.

5. Efectos previsibles de la no realización de ninguno de los procedimientos posibles.

6. Comunicación al paciente de la disposición del médico a ampliar toda la información si lo desea, y a resolver todas las dudas que tenga.

7. Comunicación al paciente de su libertad para reconsiderar en cualquier momento la decisión tomada.

Algunos autores añaden a esta lista el conocimiento de la enfermedad –etiología, curso y pronóstico– y los costes económicos de todo el proceso<sup>6</sup>.

En general existe bastante acuerdo respecto a estos diferentes ámbitos de información. Sin embargo, donde ha existido un enorme debate en EE.UU. es en lo relativo a la extensión de cada uno de los mismos; en particular, respecto a la cantidad de información relativa a los posibles riesgos, molestias y efectos secundarios de los procedimientos diagnósticos y terapéuticos.

Para intentar arrojar luz sobre esta cuestión se han utilizado en la práctica jurídica norteamericana tres criterios diferentes<sup>7,8</sup>:

1. El criterio del *médico razonable*, de la práctica profesional o de la *comunidad científica*. Este criterio, usado por primera vez en 1960 en un caso judicial americano (“Natanson frente a Kline”) estipula que la cantidad de información que debe recibir un paciente determinado viene marcada por la que un “médico razonable” revelaría en las mismas circunstancias. Un “médico razonable” es el que actúa en consonancia con la práctica habitual de la comunidad científica a la que pertenece.

El problema fundamental de este criterio es que tiene un inequívoco resabio paternalista. Por otra parte no existe garantía de que algo, por el mero hecho de ser aceptado por la comunidad científica, sea necesariamente ética o legalmente aceptable. No obstante estas críticas y a pesar de la irrupción de nuevos criterios que se verán a continuación, todavía en 1987 era un argumento ampliamente usado en el sistema judicial americano: 26 Estados lo seguían utilizando<sup>9</sup>.

2. El criterio de la *persona razonable*. El anterior criterio es propio de la década de los sesenta, sin embargo en 1969, otro caso judicial americano (“Berkey frente a Anderson”) introduce un nuevo argumento que sería sucesivamente recogido y reforzado en otros casos importantes de la década de los setenta (“Canterbury frente a Spence”, “Cobbs frente a Grant” y “Wilkinson frente a Vesey”). El razonamiento es el siguiente: si resulta que la persona apropiada para aceptar o rechazar un determinado procedimiento es el propio paciente, entonces la información que el médico tendrá que revelar a éste vendrá determinada por lo que una hipotética “persona razonable” desearía conocer en las mismas circunstancias en que se encuentra el paciente, y no por lo que otro hipotético “médico razonable” consideraría adecuado revelar. La dificultad estriba obviamente en determinar qué se entiende por “persona razonable”<sup>8</sup>.

Una forma de rodear el problema es decir que una “persona razonable” desearía conocer los “riesgos materiales”. Un riesgo es “material” si es muy grave aunque sea muy poco frecuente, o si es muy frecuente aunque sea muy poco grave.

Es decir: *materialidad = gravedad × incidencia*<sup>9</sup>.

Evidentemente, cuanto mayor sea la “materialidad” del riesgo mayor fuerza tendrá la suposición de que una “persona razonable” desearía conocerlo, y por lo tanto mayor es la obligación de comunicárselo al paciente. No es fácil, de to-

das formas, determinar qué debe entenderse por “gravedad”. Quizás es un término con unas grandes resonancias forenses. Probablemente podría calificarse de grave al riesgo que implica la posibilidad de dañar irreversiblemente el desarrollo y/o el funcionamiento normal de las capacidades biológicas, psicológicas o sociales del individuo. Se incluirían como riesgos de gravedad máxima, por ejemplo los de muerte, ceguera, hemiplejía, paraplejía, tetraplejía, pérdida de actividad sexual o mutilaciones físicas o funcionales.

Aunque el “cálculo de la materialidad” puede ser útil, dista mucho de resolver todos los problemas, pues si bien ayuda en los casos en los que la “materialidad” es elevada, permanece la cuestión del nivel más bajo de “materialidad” a partir del cual una “persona razonable” no consideraría estrictamente necesario conocer el riesgo. No obstante esta objeción importante, este criterio parece más respetuoso con la autonomía del paciente, y sin embargo en 1987 tan sólo 16 Estados lo habían introducido plenamente<sup>9</sup>.

3. El criterio *subjetivo*. En Oklahoma, el único Estado de la Unión que usa el criterio “subjetivo”, el juez pregunta al paciente demandante qué riesgos y qué complicaciones desearía haber conocido. Tal pregunta se basa en la convicción de que “cada paciente tiene unas necesidades distintas de información, especialmente los que tienen creencias idiosincráticas o peculiares, un comportamiento sanitario fuera de lo habitual, una historia personal o familiar característica o alguna otra circunstancia de este tipo”<sup>8</sup>. Este criterio implica, por lo tanto, que si un médico tiene fuertes razones para creer que su paciente desearía conocer un determinado riesgo, aunque él personalmente no lo considere de importancia “material”, tiene la obligación de comunicárselo. Este criterio ha sido mayoritariamente rechazado tanto desde el punto de vista legal como ético porque coloca a los médicos en una situación de desprotección evidente, y les obliga a someter a sus pacientes a la lectura de interminables listas de riesgos posible, antes de obtener su consentimiento, para intentar protegerse.

#### *Información con calidad suficiente: el problema de la comprensibilidad*

Como ya se ha afirmado, lo que durante las décadas de los sesenta y setenta preocupó mayoritariamente a los que reflexionaban sobre el consentimiento informado era el de la cantidad de información a proporcionar a los enfermos. Sin embargo, el progresivo descubrimiento de que la mayoría de los pacientes comprendían deficien-

temente la información tanto verbal como escrita<sup>10</sup> supuso que en la década de los ochenta haya cobrado relevancia el problema de la dificultad excesiva de comprensión de las informaciones suministradas<sup>11-13</sup>.

La cuestión es compleja, y en ella pueden considerarse dos aspectos diferentes. Existe una *dificultad objetiva* de la información transmitida por su tecnicismo, el uso de expresiones largas y enrevesadas, etc. Esta dificultad se origina en el médico, en cuanto emisor de esa información. Pero también existe una *dificultad subjetiva*, que radica en el paciente como receptor, en sus mayores o menores capacidades psicológicas para comprender lo que se le está diciendo. De aquí que este segundo tipo de dificultad esté intrínsecamente unido a uno de los requisitos de la teoría del consentimiento informado que se verá a continuación, el de la *competencia*.

La exigencia que se le plantea al médico es la de reducir al mínimo posible ambos tipos de dificultades: comunicándose con sus enfermos de la forma más sencilla y clara posible y tratando de optimizar al máximo su competencia para tomar cada decisión.

En el caso español, el artículo 10 de la Ley General de Sanidad, de nuevo en su punto 5, es bien claro al respecto cuando dice que el paciente tiene derecho a que se le proporcione información “en términos comprensibles”, tanto verbal como escrita.

#### *Competencia*

El concepto de competencia/incompetencia es quizás uno de los más complejos y nucleares de la teoría del consentimiento informado, e irrumpió en la medicina mediante el encuentro entre la psiquiatría y el Derecho. Según la teoría del consentimiento sólo los pacientes competentes tiene el derecho –ético y legal– de aceptar o rechazar un procedimiento propuesto, o sea, de otorgar o no el consentimiento. A los pacientes incompetentes se les niega ese derecho y por tanto son otros los que toman las decisiones en su lugar. En nuestro país, en virtud del punto 6, apartado b, del artículo 10 de la Ley General de Sanidad, son los familiares o personas allegadas los que ejercen esta función. La competencia podría definirse como la capacidad del paciente para comprender la situación a la que se enfrenta, los valores que están en juego y los cursos de acción posibles con las consecuencias previsibles de cada uno de ellos, para a continuación tomar, expresar y defender una decisión que sea coherente con su propia escala de valores. La de-

terminación de la competencia no es una tarea sencilla. A ello se debe quizás el hecho de que a pesar de ser un elemento tan importante del consentimiento, haya sido durante mucho tiempo un área relativamente poco trabajada dentro de la bioética norteamericana.

En 1977, tres bioeticistas americanos que se han dedicado de forma intensiva a la teoría del consentimiento informado, notificaban que los criterios que estaba usando la jurisprudencia americana para establecer la competencia podían agruparse en cinco apartados<sup>14</sup>: *a)* comunicación de decisiones; *b)* dar un motivo razonable (sensato) para estas decisiones; *c)* manifestar decisiones basadas en motivos racionales (no debidos a enfermedad); *d)* capacidad para comprender el proceso de toma de decisiones, y *e)* comprensión real y completa de ese proceso.

Posteriormente en 1982 la President's Commission desechó explícitamente el tercer criterio de la lista anterior y propuso<sup>3</sup>: *a)* la posesión del paciente de una escala de valores en que fundamentar sus decisiones; *b)* capacidad para comprender y comunicar información, y *c)* capacidad para razonar y discutir acerca de la propia opción.

El rechazo del tercer criterio no era arbitrario. Se debió a que la postura acerca de los criterios de competencia que triunfó en la bioética americana en aquellos años no era la que se fijaba en el tipo de decisión final que toma el enfermo, si ésta parece o no razonable. La postura triunfante se fijaba más en la estructura del procedimiento de toma de decisiones y de argumentación que utiliza el paciente, que en los contenidos concretos de ese proceso.

Sin embargo en 1985, James F. Drane, en un artículo del Hastings Center Report que se ha convertido en un clásico, describía un método de evaluación de la competencia que introducía un matiz importante<sup>15</sup>. Este método también valoraba el *procedimiento*, pero además las consecuencias de la acción elegida finalmente por el sujeto. Así, en la medida en que estas consecuencias son progresivamente más graves, Drane exige en su método un mayor grado de capacidad y una mayor perfección en cuanto al procedimiento para que la decisión final resulte aceptable, *competente*\*

Tanto los criterios de la President's Commission como las sugerencias de Drane han sido bastante bien acogidos en EE.UU. De hecho, Allen E. Buchanan y Dan W. Brock, dos filósofos

que formaron parte de la President's Commission y que se han dedicado de forma intensiva al problema de la competencia e incompetencia de los pacientes, han recogido estas líneas maestras en una obra, "Deciding for others. The ethics of surrogate decision making", que ya resulta un punto de referencia imprescindible a la hora de ahondar en los problemas de la toma de decisiones en el caso de pacientes incompetentes<sup>16</sup>.

Sin embargo, quedan por resolver cuestiones muy importantes. La más urgente es la elaboración de protocolos de determinación de la competencia de los enfermos que puedan ser aplicados en la práctica clínica diaria. Los trabajos más recientes de la literatura sobre competencia parecen caminar en esta dirección<sup>17,18</sup>.

### *Validez y autenticidad*

Estos requisitos no se encuentran en la definición de consentimiento informado del Colegio de Médicos Americano porque son una adquisición más tardía. En realidad están estrechamente ligadas al concepto de competencia; más bien son unas exigencias subsidiarias de ésta. El concepto de validez tiene mucho que ver con la *intencionalidad* de las acciones, la cual está muy condicionada por el estado anímico del sujeto<sup>19</sup>. Por ejemplo, una decisión tomada durante un ataque de ira puede no ser *válida* aunque el sujeto la adopte voluntaria, informada y competentemente, porque seguramente no refleja de forma adecuada sus deseos.

La *autenticidad* por su parte tiene que ver con las escalas de valores. Una decisión tomada por un sujeto voluntario, informado y competente pero que va en contra de la escala de valores que esta persona ha defendido a lo largo de toda su vida, puede no ser en realidad auténtica.

### *Las excepciones*

Todo sistema de principios y normas éticos se acompaña siempre de una lista de excepciones. Los principios y normas definen *lo bueno*, mientras que las excepciones marcan lo que se considera *mal menor*. El peligro de las excepciones es el de llegar a erigirse ellas mismas en *norma*, por eso para que resulten aceptables deben ser siempre rigurosamente justificadas<sup>20</sup>.

Las excepciones clásicas a la teoría del consentimiento, es decir, aquellas situaciones en las que se actúa sin cumplir necesariamente con sus exigencias son 6: grave peligro para la salud pública; situación de urgencia; incompetencia del enfermo; privilegio terapéutico; imperativo legal,

\*Una traducción y adaptación de la Escala de Drane al castellano puede encontrarse en Gracia D. Procedimientos de decisión en ética clínica. Madrid: Eudema, 1991; 149.

y rechazo explícito de toda la información por el paciente.

En la legislación española las excepciones al derecho al consentimiento informado se recogen en el punto 6 del artículo 10 de la Ley General de Sanidad. En él se especifican tan sólo cuatro supuestos, dejando fuera el privilegio terapéutico y el rechazo de la información por parte del enfermo. Además, la excepción por *urgencia* está redactada de tal forma que no es difícil realizar una interpretación muy paternalista de ella.

El *privilegio terapéutico*, a pesar de no haber sido reconocido por la legislación española, es una excepción que con frecuencia es usada en nuestro país como argumento para justificar la escasa información que a veces se proporciona a los pacientes. Es quizás un buen prototipo de cómo una *excepción* tiende a convertirse en *norma*. Esta excepción fue introducida en la jurisprudencia americana por el juez del caso “Canterbury frente a Spence” (1972). En el dictamen de este caso se decía<sup>21</sup>:

“El médico tiene un *privilegio* terapéutico que le capacita para ocultarle información (al paciente) respecto a los riesgos del procedimiento al que va a ser sometido, en el caso que fuera evidente que un reconocimiento médico por un profesional juicioso demostrara que tal revelación supondría una grave amenaza para la integridad psicológica del paciente”.

Las situaciones clínicas en que puede realmente ser aplicada esta excepción no son tan frecuentes como puede pensarse; por tanto deben ser rigurosamente justificadas. Los estudios que se han realizado acerca del grado de ansiedad de los pacientes a los que se les suministra información de forma amplia o no evidencian diferencias significativas respecto a otras situaciones habituales de la vida de las personas, o no son concluyentes<sup>10,22-24</sup>.

### **Maximizando el consentimiento informado**

Se decía en las líneas anteriores que las exigencias legales deben ser consideradas como mínimos éticos. Pues bien, la totalidad de los elementos de la teoría del consentimiento informado que antes se han expuesto a partir de la definición ofrecida en el Manual de Ética del Colegio de Médicos Americano, en EE.UU. son también un requisito legal. Quien los cumpla satisface tanto los mínimos éticos como los máximos legales. Sin embargo, también se ha dicho antes que éticamente debería quizá tenderse a más.

Una comprensión del consentimiento informado superadora del minimismo ético trata de introducir aquél de manera profunda en la propia concepción de la relación médico-enfermo. Tratará así de que el paciente no sólo “consienta” sino que colabore activamente con el médico en la evaluación de las opciones diagnósticas y terapéuticas, y en el proceso de toma de decisiones al respecto.

Siguiendo esta línea de pensamiento\*\* tres bioeticistas americanos, Paul S. Appelbaum (psiquiatra), Charles W. Lidz (sociólogo) y Alan Meisel (profesor de Derecho), han propuesto en una obra ya clásica de la teoría del consentimiento informado, sendos modelos de obtención del mismo<sup>25</sup>.

El primer modelo, que puede traducirse como *puntual (event model)*, entiende la toma de decisiones como un acto aislado que tiene lugar en un período de tiempo limitado, generalmente inmediatamente antes de la realización del procedimiento propuesto. En él se enfatiza sobre todo la comunicación de una cierta cantidad de información al paciente en ese momento, lo que por lo general hace que se satisfagan los requisitos éticos mínimos y por tanto legales (suponiendo, claro, que el paciente sea libre y competente para decidir). Sin embargo, la preocupación por el grado de comprensión de esa información es frecuentemente escasa, con lo que el riesgo de que este modelo caiga “bajo mínimos” es elevada. Los formularios escritos de consentimiento, tal y como habitualmente se utilizan, pueden ser considerados como el símbolo central de este modelo.

El segundo modelo, al que se traducirá como *progresivo (process model)*, se basa en la participación activa del paciente en el proceso de toma de decisiones, e introduce la obtención del consentimiento informado como un requisito imprescindible de todos y cada uno de los pasos sucesivos que componen globalmente este proceso. La toma de decisiones en el seno de la relación médico-enfermo no es un hecho aislado sino un proceso continuo que comienza en el mismo instante en que el paciente acude al médico, y que finaliza cuando la salud objetiva y subjetiva ha sido recobrada. Aunque es cierto que existen momentos concretos dentro de esta línea continua en los que se toman decisiones más relevantes, la obtención del consentimiento informado no se limita tan sólo a estos instantes.

\*\*Una versión resumida puede encontrarse en Lidz CW, Appelbaum PS, Meisel A. Two models of implementing informed Consent. Arch Intern Med 1988; 148: 1.385-1.389.



Aunque el modelo *progresivo* es más deseable, presenta evidentemente límites en su aplicación. Sólo puede ser desarrollado por aquellos médicos que tengan un contacto directo y prolongado con sus pacientes. Determinadas especialidades médicas, como la radiología por ejemplo, se verán obligadas a ceñirse al modelo puntual, aunque evidentemente traten de maximizarlo en lo posible. Por otra parte, aun en el seno del modelo *progresivo* existirán momentos concretos, cuando haya que tomar decisiones muy importantes, en los que las pautas de comportamiento estarán muy cercanas al modelo *puntual*, incluso utilizando formularios escritos.

## BIBLIOGRAFÍA

1. Centro de Investigaciones Sociológicas. La asistencia sanitaria en España I y II. Revista Española de Investigaciones Sociológicas 1991; 53: 267-303. 54: 245-267.
2. Ad hoc Committee on Medical Ethics, American College of Physicians. American College of Physicians Ethics Manual. Ann Intern Med 1984; 101: 129-137, 263-274.
3. President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research: Making Health Care Decisions. Washington DC: Government Printing Office, 1982; 1.
4. Meisel A, Roth HL, Lidz CW. Toward a model of the legal doctrine of informed consent. Am J Psychiatry 1977; 134: 285-289.
5. Alfidi RJ. Controversy, alternatives and decisions complying with the legal doctrine of informed consent. Radiology 1975; 114: 231-234.
6. Morgan LW, Schwab IR. Informed consent in senile cataract extraction. Arch Ophthalmol 1986; 104: 42-45.
7. Gracia D. Fundamentos de Bioética. Madrid: Eudema, 1989; 165-173.
8. Beauchamp TL, McCullough LB. Ética Médica. Barcelona: Labor, 1987; 76-79.
9. Reuter, SR. An overview of informed consent for radiologists. AJR 1987; 148: 219-227.
10. Meisel A, Roth LH. What we do and do not know about informed consent. JAMA 1981; 246: 2.473-2.477.
11. Bergler JH, Pennington AC, Metcalfe M, Freis ED. Informed consent: How much does the patient understand? Clin Pharmacol Ther 1980; 27: 435-440.
12. Riecken HW, Ravich R. Informed consent to biomedical research in Veterans Administration Hospitals. JAMA 1982; 248: 344-348.
13. Byrne DJ, Napier A, Cuschieri A. How informed is signed consent? Br Med J 1988; 296: 839-840.
14. Roth LH, Meisel A, Lidz CW. Tests of competency to consent to treatment. Am J Psychiatry 1977; 134: 279-284.
15. Drane JF. The many faces of competency. Hastings Cent Rep 1985; 4: 17-21.
16. Buchanan AE, Brock DW. Deciding for others. The ethics of surrogate decisions. Cambridge: Cambridge University Press, 1989.
17. Kaplan K, Price M. The clinician's role in competency evaluations. Gen Hosp Psychiatry 1989; 11: 397-403.
18. Draper RJ, Dawson D. Competence to consent to treatment: a Guide for the Psychiatrist. Can J Psychiatry 1990; 35: 285-289.
19. Couceiro Vidal A. La información al paciente. JANO 1991; (extra III): 77-82.
20. Gracia D. Procedimientos de decisión en ética clínica. Madrid: Eudema, 1991; 134-136.
21. Brahams D. Doctor's duty to inform patient of substantial or special risks when offering treatment. Lancet 1985; 1: 528-530.
22. Spring DB, Winfield AC, Friedland GW, et al. Written informed consent for IV contrast-enhanced radiography: Patients' attitudes and common limitations. AJR 1988, 151: 1.243-1.245.
23. Simes RJ. Randomised comparison of procedures for obtaining informed consent in clinical trials of treatment for cancer. Br Med J 1986; 293: 1.065-1.068.
24. Myers MG, Cairns JA, Singer J. The consent form as a possible cause of side effects. Clin Pharmacol Ther 1987; 42: 250-253.
25. Appelbaum PS, Lidz CW, Meisel A. Informed Consent, Legal Theory and Clinical Practice. Oxford University Press, 1987; 151-174.

---

# Legibilidad de los formularios escritos de consentimiento informado

---

Pablo Simón Lorda<sup>a</sup>, Inés M.<sup>a</sup> Barrio Cantalejo<sup>b</sup> y Luis Concheiro Carro<sup>c</sup>

<sup>a</sup>Especialista en Medicina Familiar y Comunitaria. Magister en Bioética. Instituto de Medicina Legal. Universidade de Santiago de Compostela. Galicia. <sup>b</sup>DUE. Experta en Bioética. Profesora de "Ética y Legislación de Enfermería". Escuela Universitaria de Enfermería - La Paz. Universidad Autónoma de Madrid. <sup>c</sup>Catedrático de Medicina Legal. Instituto de Medicina Legal. Universidade de Santiago de Compostela. Galicia.

El uso del lenguaje escrito como forma de proporcionar información a los pacientes es algo muy frecuente en el mundo sanitario. En los últimos años ha cobrado todavía mayor relevancia a medida que el consentimiento informado ha sido reconocido como la parte esencial de la relación médico-paciente. El consentimiento informado no es sólo un requisito legal sino, sobre todo, una exigencia ética del sanitario, en cuyo cumplimiento los formularios escritos desempeñan un papel de apoyo insustituible, siempre y cuando sean confeccionados y usados correctamente<sup>1</sup>. Las funciones básicas de los formularios escritos de consentimiento informado son, por orden de importancia, las tres siguientes: orientar el proceso de información y de debate sobre la medida diagnóstica o terapéutica propuesta entre sanitario y paciente, proceso que culmina cuando el paciente autónomamente decide si acepta o rechaza tal medida; servir de registro de que ese proceso se ha llevado a cabo y, por último, facilitar –que no garantizar– la protección legal del sanitario.

Para cumplir adecuadamente su cometido, los formularios escritos tienen que ser específicos para cada procedimiento diagnóstico y terapéutico, y constar de dos partes diferenciadas, una destinada a proporcionar al paciente una cierta cantidad de información básica, y otra cuyo objetivo es registrar que efectivamente el consentimiento fue emitido. La información facilitada al enfermo tiene que ser suficiente en cantidad y calidad. La variable fundamental para medir esta última es su grado de legibilidad. No cumple ninguna de las tres funciones descritas aquel formulario que esté escrito en un lenguaje difícil de entender por el paciente. Por eso resulta imprescindible contar con instrumentos que nos permitan evaluar objetiva-

mente el grado de legibilidad de los formularios. En el mundo anglosajón existen varios instrumentos de análisis de la legibilidad que han sido aplicados a los formularios de consentimiento –sobre todo para la participación en investigación– u otros materiales escritos dirigidos a los pacientes<sup>2</sup>: la fórmula de Flesch<sup>3</sup>, la escala de Fry<sup>4</sup>, el "Gunning Fog Test"<sup>5</sup>, y otros. Algunos de ellos existen ya en versiones informatizadas, como una herramienta más de los programas de tratamiento de textos. La más empleada de todas ha sido, sin duda, la fórmula de legibilidad de Flesch. En nuestro país se ha realizado ya un estudio piloto de validación de dicha fórmula para su aplicación al análisis de formularios escritos de consentimiento informado<sup>6</sup>, pero no se disponía hasta hace bien poco de una versión informatizada de dicha técnica, lo que limitaba su utilidad práctica dado lo engoroso de su aplicación manual. La recién comercializada versión 6.1 del programa Word Perfect incorpora la primera versión en castellano de la utilidad Grammatik 6.0, que contiene una forma actualizada de la técnica de Flesch, el índice de legibilidad de Flesch-Kincaid (*Flesch-Kincaid Grade Level*), además de otras mediciones útiles para el análisis de la legibilidad, como los usos de la voz pasiva, la complejidad de las oraciones o la complejidad del vocabulario<sup>7</sup>.

Los objetivos del presente trabajo son los siguientes: *a)* analizar el comportamiento de las aplicaciones de Grammatik sobre la legibilidad ante textos de diferente tipo y presumiblemente variada dificultad; *b)* evaluar el calibrado de dichas aplicaciones realizado por la empresa comercializadora del programa y, eventualmente, reajustar si se precisa el sistema de medición, o proponer uno nuevo complementario; y *c)* evaluar, con estos instrumentos, la legibilidad de un grupo de formularios de consentimiento informado usados en nuestro país.

## Material y métodos

### Textos analizados

Se escogieron diferentes textos para ser analizados. Pueden ser agrupados en tres clases:

a) *Textos generales*. Se recogieron 10 textos pertenecientes a diferentes publicaciones o libros accesibles al ciudadano medio, pero con grados de legibilidad presumiblemente muy diferentes. Muchos de ellos pertenecen a los mismos libros o publicaciones utilizados en el estudio piloto de validación de la técnica de Flesch publicado en 1993<sup>6</sup>. En la medida de lo posible se procuró que tuvieran en torno a 500 palabras, pues la mayoría de los formularios de consentimiento informado en nuestro país tienen 400 y 500 palabras. Los textos fueron los siguientes: el artículo 10 de la Ley General de Sanidad de 1986, que contiene la Carta de Derechos de los pacientes; la primera página de los periódicos *El País*, *El Mundo* y *Diario 16*; los primeros párrafos de *El Quijote* de Cervantes<sup>8</sup>, del diálogo platónico *La República*<sup>9</sup>, y del libro de Génesis<sup>10</sup>; un texto de literatura infantil titulado *Los Mifenses* de Rocio de Terán, pensado para niños de más de 9 años<sup>11</sup>, y dos cómics, uno de adultos –*El Víbora*– y otro infantil –*Mortadelo y Filemón*–.

b) *Publicaciones periódicas*. Algunas publicaciones periódicas contienen textos muy variados en estilo y dificultad, por lo que resulta difícil evaluar en conjunto su grado de legibilidad. Para poder llevar a cabo una estimación general de sus medidas de legibilidad con una confianza del 90% y una precisión de  $\pm 10$  se estimó que debían escogerse 17 muestras de  $\pm 100$  palabras de cada una de las publicaciones. Se escogieron 8 publicaciones periódicas ya analizadas en el estudio piloto<sup>6</sup>. El motivo de que las muestras sean de unas 100 palabras estriba en que es la unidad clásica de medida usada por la técnica de Flesch. Con el objetivo de simplificar el estudio, para el cálculo del tamaño de la muestra se estimó como desviación estándar un valor de 25 (100/4) para todos los casos, en lugar de aplicar en cada uno los valores obtenidos en el estudio piloto<sup>6</sup>. La selección de las páginas donde se debía escoger cada muestra se realizó aplicando una tabla de números aleatorios. La muestra empezaba a contarse en cada página a partir del primer punto y aparte que apareciera. Las 8 publicaciones así analizadas fueron las siguientes: *MEDICINA CLÍNICA*<sup>12</sup>, *Cirugía Española*<sup>13</sup>, *The Lancet* (ed. esp.)<sup>14</sup>, *Jano*<sup>15</sup>, *Semana*<sup>16</sup>, *Muy Interesante*<sup>17</sup>, *Pronto*<sup>18</sup> y *El País Semanal*<sup>19</sup>. Para que

los resultados fueran más comparables con los del estudio piloto, el número de *Jano* analizado pertenece a la serie antigua, no a la del nuevo formato de la revista puesto en marcha en 1995.

c) *Formularios de consentimiento informado*. Se analizaron 16 formularios de consentimiento informado que, o bien son utilizados –o lo han sido hasta fechas recientes– en distintos centros sanitarios españoles, o bien han sido propuestos para su uso en la bibliografía médica de nuestro país. Los formularios extraídos de la bibliografía son los propuestos para las unidades de cuidados intensivos<sup>20</sup>, el electroshock<sup>21</sup>, el test serológico del virus de la inmunodeficiencia humana<sup>22</sup>, la angiografía y un formulario semiespecífico para cirugía<sup>6</sup>. Dos de los formularios, los de anestesia general y anestesia locorregional, han sido oficialmente recomendados a los centros sanitarios dependientes del Instituto Nacional de la Salud (INSALUD) por la Dirección General del citado organismo en junio de 1994. Pero no sólo son utilizados en dichos centros, sino también en otros dependientes de Administraciones sanitarias autonómicas como, por ejemplo, en hospitales del Servicio Galego de Saúde (SERGAS). El resto de los formularios han sido elaborados en distintos centros sanitarios y se han escogido, de entre los muchos disponibles, como algunos de los más interesantes y de redacción más cuidada. Estos formularios son los siguientes: trasplante renal (Hospital Infantil La Paz, Madrid); transfusión sanguínea (Hospital General de Castellón); biopsia hepática y laparoscopia –lo llamaremos “biopsia hepática 1”– (Hospital 12 de Octubre, Madrid); biopsia hepática –lo llamaremos “biopsia hepática 2”– (Hospital Virgen de la Salud, Toledo); biopsia hepática –lo llamaremos “biopsia hepática 3”– (Consorti Hospitalari del Parc Taulí, Sabadell, Barcelona); laparoscopia (Hospital Virgen de la Salud, Toledo); endoscopia digestiva (Hospital Universitari Sant Joan, Reus, Tarragona); operación de cataratas (Hospital General Yagüe, Burgos), y angiografía fluoresceínica (Hospital General Yagüe, Burgos).

### Instrumentos de análisis

Se utilizó el programa Word Perfect 6.1 en versión para Windows, instalado en un ordenador PC compatible HP Vectra VL2 4/50. Este programa incorpora la utilidad Grammatik 6.0. Dicha utilidad permite realizar diversos tipos de análisis sintácticos y gramaticales al texto al que se le aplica y con variados grados de profundidad. En este caso se empleó la “verificación exhaustiva” que es la forma más completa de evaluación que

puede realizarse, y no se modificaron los valores que el programa asigna por defecto a los criterios de revisión. Una vez dentro de Grammatik los comandos *Ver/Estadísticas* permiten visualizar un primer nivel de análisis de la legibilidad del texto, consistente en un recuento de una serie de parámetros: número de palabras, número de oraciones, número de párrafos, palabras cortas, palabras largas o número medio de palabras por frase. De todos ellos, los únicos que se reflejarán en los resultados de este estudio son los del número de palabras y oraciones de cada texto, pero sólo con intención descriptiva. Desde esta pantalla el uso del comando *Legibilidad* permite visualizar los resultados de cuatro índices distintos de medición de la legibilidad<sup>23</sup>. De los cuatro, en el presente estudio, sólo se emplearán dos de ellos:

*Índice de legibilidad de Flesch-Kincaid.* Aplica la técnica de Flesch, basada en la relación existente entre la dificultad del texto y la longitud de las palabras y de las frases. La puntuación obtenida puede oscilar entre 0 (legibilidad mínima) y 100 (legibilidad máxima). Es importante hacer notar que no es posible obtener puntuaciones negativas, y que éstas se asimilan siempre a un valor 0, lo que establece una diferencia importante respecto a los resultados del estudio piloto<sup>6</sup>.

*Índice de complejidad de oraciones.* Al parecer –porque la información del manual del programa no se extiende excesivamente en explicarlo–, realiza un análisis de la estructura sintáctica de cada una de las oraciones del texto. Evidentemente cuantas más oraciones coordinadas y subordinadas, y cuantas menos oraciones simples tenga el texto, más difícil será de leer. Toma valores entre 0 (complejidad mínima y, por tanto, legibilidad máxima) y 100 (complejidad máxima y, por tanto, legibilidad mínima). Su sistema de puntuación es, por tanto, el inverso del que tiene el índice de Flesch-Kincaid.

El motivo de no utilizar el índice denominado *complejidad del vocabulario* estriba en que, para analizar la dificultad de las palabras empleadas en el texto, las compara con el diccionario acoplado a Grammatik, que contiene el repertorio de las palabras de uso más común. Su sistema de puntuación va de 0 (“muy fácil”) a 100 (“muy difícil”). Cualquier palabra, sea del tipo que sea, que no esté en dicho diccionario, eleva considerablemente el nivel de dificultad terminológica del texto. La mayoría de los textos que se han utilizado en este trabajo obtuvieron puntuaciones entre 70 y 80. Es por tanto un índice con una elevada sensibilidad, pero con una especificidad

prácticamente nula. El otro índice, *usos de la voz pasiva*, no fue investigado.

El índice de Flesch-Kincaid y el índice de complejidad de oraciones están evidentemente relacionados de forma estrecha, pero no son idénticos. Es obvio que la tendencia general será que, cuanto más bajo sea el primero, más alto será el segundo, porque ello querrá decir que el texto tiene frases muy largas (bajo índice de Flesch) y, por tanto, esas frases probablemente serán casi todas compuestas (alto índice de complejidad de oraciones). Pero ello no es necesariamente siempre así, pues incluso frases medianamente largas (valor medio del índice de Flesch), pueden en realidad estar compuestas de varias frases pequeñas (alto índice de complejidad de oraciones). Ambos índices realizan, pues, análisis distintos de la legibilidad del texto, pero complementarios. Es por ello que se decidió integrarlos en uno solo, al que se denominó *legibilidad integrada* (LEGIN), en la convicción de que se aumentará la sensibilidad y la especificidad de la medición final de la legibilidad. Este índice se obtiene mediante una sencilla fórmula:  $LEGIN = 100 + \text{índice de Flesch} - \text{índice de complejidad de oraciones}$ . El LEGIN toma valores entre 0 (legibilidad mínima) y 200 (legibilidad máxima).

Es fundamental tener en cuenta que el análisis de la legibilidad que realiza Grammatik no tiene en cuenta encabezados, notas a pie de página, y en general, cualquier texto que posea un “formato especial”. Sin embargo ni el comando *Ayuda* del programa ni el manual de uso dan instrucciones sobre lo que deben considerarse “formatos especiales” ni orientan sobre la forma en la que deben escribirse los textos antes de ser analizados con Grammatik. En la tabla I se encuentran algunas de estas recomendaciones, obtenidas empíricamente durante la realización de este trabajo. Curiosamente algunas son muy similares a las recomendaciones extraídas del estudio piloto de validación de la técnica de Flesch<sup>6</sup>. Son muy importantes, porque si no se tienen en cuenta Grammatik excluirá del análisis porciones extensas del texto, con lo que los resultados obtenidos serán muy poco fiables.

### Metodología

Se introdujeron en el programa Word Perfect, en ficheros independientes, todos y cada uno de los textos anteriormente señalados, teniendo en cuenta las recomendaciones expresadas al respecto en la tabla I. A continuación se les aplicó la utilidad Grammatik y se obtuvieron 4 parámetros: número de palabras, número de frases, puntua-

TABLA I  
NORMAS PRÁCTICAS PARA EL USO CORRECTO DE LOS SISTEMAS DE MEDICIÓN  
DE LA LEGIBILIDAD DE GRAMMATIK

<p>Grammatik cuenta como oración entera la que se encuentra entre un punto y otro –sean punto y aparte o punto y seguido–. Por tanto:</p> <ol style="list-style-type: none"> <li>1. Comenzar una frase enumerándola con expresiones como " 1. La anestesia (...)", " A. La anestesia (...)" o " a. La anestesia (...)" hará que el programa no la cuente como una frase sino como dos, lo que distorsionará a la baja la dificultad del texto</li> <li>2. Igual sucederá con expresiones como " Dr. (...)", " Dra. (...)", " Vd. ". Es por ello conveniente utilizar el punto exclusivamente para separar frases. Sin embargo no parece haber problema al incluirlo en cifras numéricas, " 2.000", " 14.567", que son leídas sin dificultad por Grammatik</li> <li>3. Olvidarse de poner el punto al final de una oración es arriesgarse a que Grammatik la una con la oración siguiente como una única frase en conjunto, lo que también distorsionará, pero esta vez al alza, la dificultad del texto</li> </ol> <p>Grammatik considera formato especial a las oraciones comenzadas por guiones o asteriscos, y por tanto no serán tenidas en cuenta por el programa. Es conveniente también evitar encabezados como " a)" o " 1)". Expresiones como " Sr./Sra.", " y/o"</p>	<p>dentro de un texto pueden resultar difíciles de captar por Grammatik</p> <p>No parece haber problema con las frases entre interrogantes o admiraciones, o con negrita o cursiva, pero las frases entrecorridas pueden resultar conflictivas. Las frases entre paréntesis o entre guiones es mejor separarlas mediante comas</p> <p>Los dos puntos resultan difíciles de interpretar por el programa y pueden ser fuente de error, por lo que es conveniente no usarlos. No parece haber problema con el punto y coma</p> <p>Las líneas de puntos y/o rayas, los bordes, las cajas de texto, las tramas de fondo, y todo aquello que sugiera un formato especial debe ser eliminado antes de utilizar Grammatik</p> <p>En general, deben sustraerse al análisis aquellas partes del texto que no aportan nada relevante al mismo desde el punto de vista de su contenido informativo. Por ejemplo, mantener un título como " A. Información." supone incrementar al texto en dos frases de una largura mínima, lo que supone aumentar artificialmente la legibilidad del conjunto. Frases como " Firma del paciente.", " Firmado." o " Fecha." tienen el mismo efecto distorsionador. Es mejor suprimirlas antes del uso de Grammatik</p>
--	---

ción de Flesch y complejidad oracional. El número de palabras y frases se considerará sólo de forma descriptiva. En el caso de la *publicaciones periódicas* se obtuvo la media y desviación estándar de los tres últimos parámetros en el conjunto de las 17 muestras de  $\pm 100$  palabras, y se calculó el intervalo de confianza. Por último, se calculó el LEGIN de cada uno de los textos.

## Resultados

### *Análisis de los textos generales y las publicaciones periódicas*

En las tablas II y III se encuentran los valores obtenidos en cada uno de los 5 parámetros. Además, se señalan las abreviaturas usadas para identificar los textos en las figuras. En la figura 1 se representa el comportamiento del índice de Flesch y del índice de complejidad de oracio-

nes. En la figura 2 se exponen gráficamente los valores del LEGIN.

### *Análisis de los formularios de consentimiento informado*

En la tabla IV se encuentran los valores obtenidos en el análisis de los 16 formularios escritos de consentimiento informado, así como las abreviaturas utilizadas para identificarlos en las figuras. Las figuras 3 y 4 recogen, respectivamente, representaciones gráficas de los índices de Flesch y complejidad de oraciones, y de los valores del LEGIN.

## Discusión

La combinación del índice de Flesch y del índice de complejidad de oraciones, proporcionados por la utilidad Grammatik del programa de

TABLA II  
RESULTADOS DEL ANÁLISIS DE LOS 10 TEXTOS GENERALES

Texto	Abreviatura utilizada	N.º de palabras	N.º de frases	Flesch	Comp. or.	LEGIN
Art. 10 - L.G. Sanidad	LGS	494	21	0	59	41
<i>El País</i> , 1.ª pág.	PAIS	778	46	13	37	76
<i>El Quijote</i> , cap. 1	QUI	558	18	19	71	49
<i>El Mundo</i> , 1.ª pág.	MUN	697	49	24	22	102
<i>Diario 16</i> , 1.ª pág.	D16	353	46	29	5	124
Platón, <i>La República</i>	PLA	703	45	29	34	95
<i>Los mifenses</i> (lit. infantil)	MIF	598	58	31	11	120
<i>El Víbora</i> (cómic adultos)	VIB	513	77	43	9	134
El Génesis (Gen.1,1-2,4)	GEN	713	54	51	24	127
<i>Mortadelo y Filemón</i>	MOR	570	115	51	6	145

Comp. or.: puntuación del índice de complejidad de oraciones; Flesch: puntuación del índice de legibilidad de Flesch-Kincaid; LEGIN: puntuación del índice de legibilidad integrada.

tratamiento de textos Word Perfect 6.1 ofrece una descripción fiable –aunque siempre aproximada– de la legibilidad de un texto. El uso del índice integrado que hemos denominado LEGIN

permite obtener un valor sintético que resume adecuadamente la legibilidad general del escrito.

En el caso del índice de Flesch esta afirmación se desprende del hecho de que los valores obtenidos

TABLA III  
RESULTADOS DEL ANÁLISIS DE LAS PUBLICACIONES PERIÓDICAS

Publicación	Abreviatura utilizada	N.º de palabras	N.º de frases	Flesch	Comp. or.	LEGIN
MEDICINA CLÍNICA	MED	1.707	50	0 (0)	58,9 (11,6)	41
<i>Cirugía Española</i>	CIR	1.783	67	[-10-10] 0,7 (1,7)	[48,9-68,9] 53,6 (10,9)	47
<i>Lancet</i>	LAN	1.819	70	[-10,7-10,7] 1,7 (4,2)	[43,6-53,6] 51,1 (16,9)	50,5
<i>Jano</i>	JAN	2.062	86	[-8,3-11,7] 3,5 (5,5)	[41,1-61,1] 50,7 (18,9)	52,8
<i>Semana</i>	SEM	2.005	89	[-6,4-13,5] 14 (11,3)	[40,7-60,7] 46,6 (16,8)	67,4
<i>Muy Interesante</i>	MUY	1.875	130	[4-24] 17,7 (14)	[36,6-56,6] 34 (15,8)	83,7
<i>Pronto</i>	PRO	1.982	96	[7,7-27,7] 21 (14)	[24-44] 50,6 (17,5)	70,3
<i>El País Semanal</i>	PSE	1.789	146	[11-31] 23,2 (11,9)	[40,6-60,6] 21,6 (15,9)	101,6
				[13,2-33,2]	[11,6-31,6]	

Comp. or.: puntuación del índice de complejidad de oraciones; desviación estándar = (Y); Flesch: puntuación del índice de legibilidad de Flesch-Kincaid; intervalo de confianza = [X-Y]; LEGIN: puntuación del índice de legibilidad integrada; media aritmética = X.

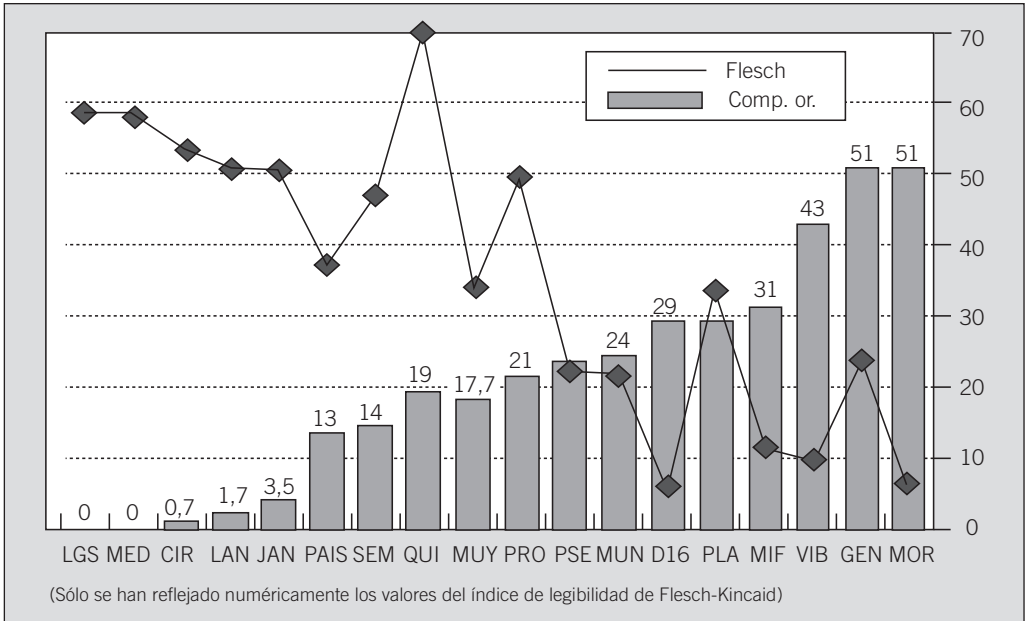


Fig. 1. Representación de los valores del índice de Flesch y del índice de complejidad oracional obtenidos al analizar los textos generales y las publicaciones periódicas; CIR: Cirugía Española; Comp. or.: puntuación del índice de complejidad de oraciones; D16: Diario 16; GEN: El Génesis; JAN: Jano; LAN: The Lancet (ed. esp.); LGS: Ley General de Sanidad; MED: MEDICINA CLÍNICA; MIF: Los Mifenses; MOR: Mortadelo y Filemón; MUN: El Mundo; MUY: Muy Interesante; PAIS: El País; PLA: La República (Platón); PRO: Pronto; PSE: El País Semanal; QUI: El Quijote; SEM: Semana; VIB: El Víbora.

nidos se correlacionan de forma adecuada con las legibilidades presumiblemente esperadas en cada caso, dado el tipo de texto o publicación y su presunta accesibilidad para el ciudadano medio. Esta gradación en los valores del índice de Flesch es correlativa de la estructura general de puntuación que proporciona Flesch para textos en inglés, así como de valores la que, para similares textos y publicaciones, se obtuvo en el estudio piloto –salvando la imposibilidad de obtener en este caso puntuaciones negativas<sup>6</sup>. En ese estudio piloto se recomendaba que los textos dirigidos al gran público tuvieran valores superiores a 0, y que las legibilidades óptimas se obtenían con puntuaciones superiores a 10. Vistos los resultados del estudio, es mejor afirmar que la puntuación de Flesch debe ser siempre superior a 10, puesto que es por encima de ese valor donde se sitúan los textos y publicaciones destinados al ciudadano medio. No parece, pues, acertada la afirmación del manual del programa que insiste en que “un índice de legibilidad entre 6 y 10 es el más eficaz para dirigirse al público en general”.

El índice de complejidad de oraciones aporta una información suplementaria muy sugerente sobre la legibilidad del texto. Como puede verse en las figuras 1 y 3 la impresión de que las puntuaciones del índice de Flesch y del índice de complejidad de oraciones tenderían a ser, en general, simétricas pero inversas parece confirmarse. Pero lo más interesante son los casos en los que esta norma genérica no se cumple, porque proporcionan una información suplementaria sobre la legibilidad de alto valor. El caso del texto de *El Quijote* es a este respecto paradigmático. Tiene un índice de Flesch de 19 –más que aceptable– pero una complejidad oracional de 71, la más alta de todo el estudio. Si se repasa la estructura de ese texto se observa que tiene un alto grado de complejidad sintáctica, con frases no muy largas, pero sí extraordinariamente complejas, con muchas frases subordinadas. Es seguro que el texto de *El Quijote* no es fácil de leer por el ciudadano medio.

Por eso, el índice LEGIN ofrece una visión integrada y sintética de la legibilidad, casi con toda seguridad de mayor sensibilidad y especificidad

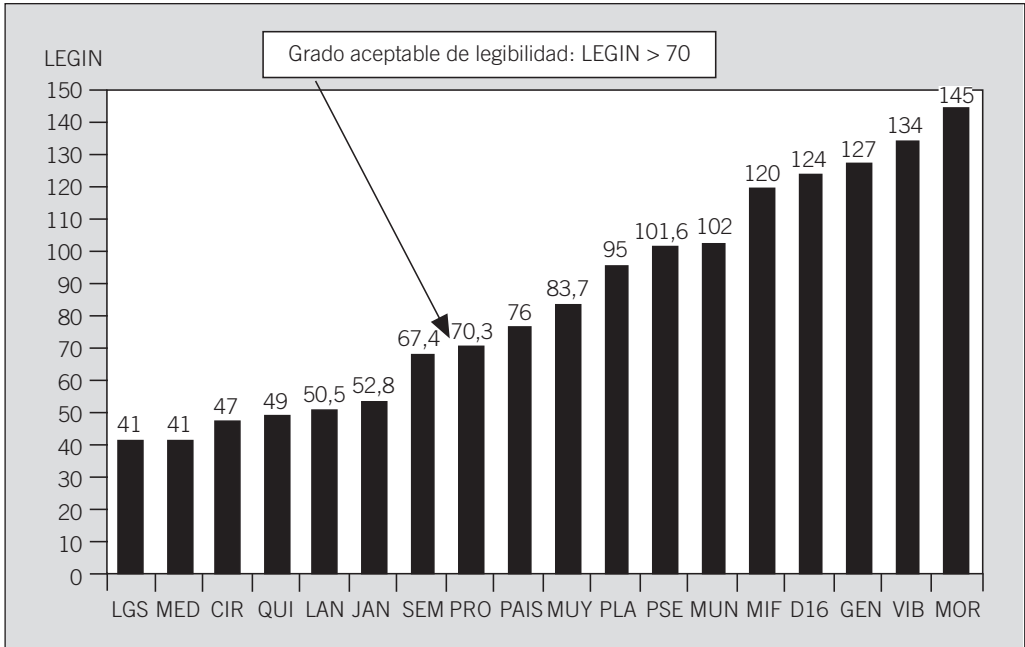


Fig. 2. Representación gráfica de los valores del LEGIN en el análisis de los textos generales y las publicaciones periódicas; CIR: Cirugía Española; D16: Diario 16; GEN: El Génesis; JAN: Jano; LAN: The Lancet (ed. esp.); LGS: Ley General de Sanidad; MED: MEDICINA CLÍNICA; MIF: Los Mifenses; MOR: Mortadelo y Filemón; MUN: El Mundo; MUY: Muy Interesante; PAÍS: El País; PLA: La República (Platón); PRO: Pronto; PSE: El País Semanal; QUI: El Quijote; SEM: Semana; VIB: El Víbora.

que la de los otros dos índices por separado. Del análisis de los valores del LEGIN en la figura 2 se desprende que hacia el valor de 70 puede establecerse la línea divisoria entre las legibilidades que son aceptables y las que no. Es decir, un texto destinado al ciudadano medio debería tener valores del LEGIN superiores a 70. Con este índice *El Quijote* queda correctamente clasificado como un texto difícil de leer, pues tiene un LEGIN de 49. Con respecto a la publicación *Semana*, que queda por debajo del valor de 70, aunque muy próxima, hay que señalar que tiene un estilo literario totalmente deplorable, con muy pocos signos de puntuación o mal colocados, frases bastante largas con un gran número de subordinadas, e incluso, frases sin sentido. De todas formas hay que tener en cuenta que las puntuaciones reales de sus índices de Flesch y de complejidad de oraciones vienen definidas por el intervalo de confianza, y no por valores absolutos.

Por tanto, en conjunto y como orientación genérica, puede decirse que un texto escrito destinado a ser leído por el público en general debe-

ría tener un índice de Flesch superior a 10, un índice de complejidad de oraciones inferior a 40 y, sobre todo, un LEGIN mayor de 70. De lo contrario las posibilidades de que sea comprendido disminuirán notablemente.

Los resultados obtenidos en el análisis de los 16 formularios escritos de consentimiento informado hablan por sí solos. Sólo 5 formularios le resultan accesibles al ciudadano medio si se tienen en cuenta los valores del índice de Flesch o los del LEGIN. Sin embargo, hasta siete muestran valores adecuados de complejidad de oraciones. Es interesante que entre los 5 formularios con LEGIN aceptable hay uno, el de la UCI, que tenía una puntuación de Flesch de 0; sin embargo, el valor de 24 que tiene su complejidad de oraciones lo vuelve fácil de leer. Por el contrario, el formulario semiespecífico de cirugía (CIR) que tenía una puntuación de Flesch de 11 cae a un LEGIN de 69 debido a una complejidad de oraciones de 42. En cualquier caso, lo más importante es señalar que es dudoso que todos los demás formularios que no tienen ni una sola puntuación acep-



TABLA IV  
RESULTADOS DEL ANÁLISIS DE LOS 16 FORMULARIOS ESCRITOS  
DE CONSENTIMIENTO INFORMADO

Formularios	Abreviatura utilizada	N.º de palabras	N.º de frases	Flesch	Comp. or.	LEGIN
Biopsia hepática 1	BH1	404	15	0	62	38
Trasplante renal	TRR	385	15	0	61	39
Transfusión	TRF	299	13	0	58	42
Anestesia locorregional	ALO	542	23	0	57	43
Laparoscopia	LAP	325	17	0	47	53
Endoscopia	END	214	11	0	47	53
Anestesia general	AGE	538	29	0	57	64
Unidad de cuidados intensivos	UCI	302	17	0	24	76
Cirugía de cataratas	OF1	594	29	1	50	51
Biopsia hepática 2	BH2	501	29	2	37	65
Angiofluoresceína	OF2	481	25	3	49	54
Electroshock	TEC	468	29	10	31	79
Biopsia hepática 3	BH3	302	23	11	22	92
Cirugía (formato semiespecífico)	CRG	443	26	11	42	69
Angiografía	ANG	519	37	21	27	94
Test del virus de la inmunodeficiencia humana	VIH	561	47	24	24	100

Comp. or.: puntuación del índice de complejidad de oraciones; Flesch: puntuación del índice de legibilidad de Flesch-Kincaid; LEGIN: puntuación del índice de legibilidad integrada.

table cumplan adecuadamente las tres funciones que les han sido señaladas. La baja legibilidad de los formularios de consentimiento informado es algo ya reiteradamente señalado por la bibliografía anglosajona, que se ha dedicado sobre todo a estudiar los formularios de consentimiento para la participación de ensayos clínicos, aunque también los formularios para procedimientos asistenciales, y ha utilizado fundamentalmente para sus análisis la fórmula de Flesch<sup>24-30</sup>.

Lograr una legibilidad aceptable en los formularios escritos de consentimiento informado es importante, sobre todo si se desea utilizarlos correctamente, esto es, pensando sobre todo en su función de guía del proceso de información y consentimiento. En tales casos el sanitario entrega al paciente el formulario con tiempo suficiente para que lo vaya leyendo, solo o, si así lo desea, junto con su familia. Simultáneamente, o más tarde según cada caso, el sanitario aclara las dudas que pueda plantear el formulario, o matiza o amplía la información contenida en el mismo. Sólo una vez que el diálogo ha concluido y el paciente ha tomado una decisión se procede a la firma del documento. Más allá de los déficit intrínsecos de comprensión que puedan radicar en el propio paciente –y que en su versión más ex-

trema pueden hacernos dudar de su competencia para tomar decisiones–, es obvio que el proceso señalado no puede realizarse si el texto escrito resulta totalmente ininteligible al paciente. En tales casos la firma del formulario se reducirá con frecuencia a un mero trámite burocrático sin ningún contenido real, más que el de ofrecer una vaga protección legal al sanitario. “Vaga” porque lo jurídicamente decisivo no es un papel firmado, sino poder demostrar que el proceso de información y consentimiento se llevó a cabo, y mal comienzo es presentar como prueba un papel que posiblemente ni el mismo juez sea capaz de entender.

Pero más allá de las cuestiones legales, lo que a la larga revela el análisis de la legibilidad de los formularios es la sensibilidad ética de quien los redacta. Esto no debe entenderse como un juicio sobre la calidad moral de los que han elaborado los formularios analizados. Antes al contrario, dichos formularios han sido con toda seguridad redactados por personas con una enorme preocupación por hacerlo lo mejor posible, y con un alto grado de seriedad moral. Lo que sucede es que el personal sanitario, como personal técnico muy especializado, tiene grandes dificultades para explicarse con palabras sencillas, comprensibles

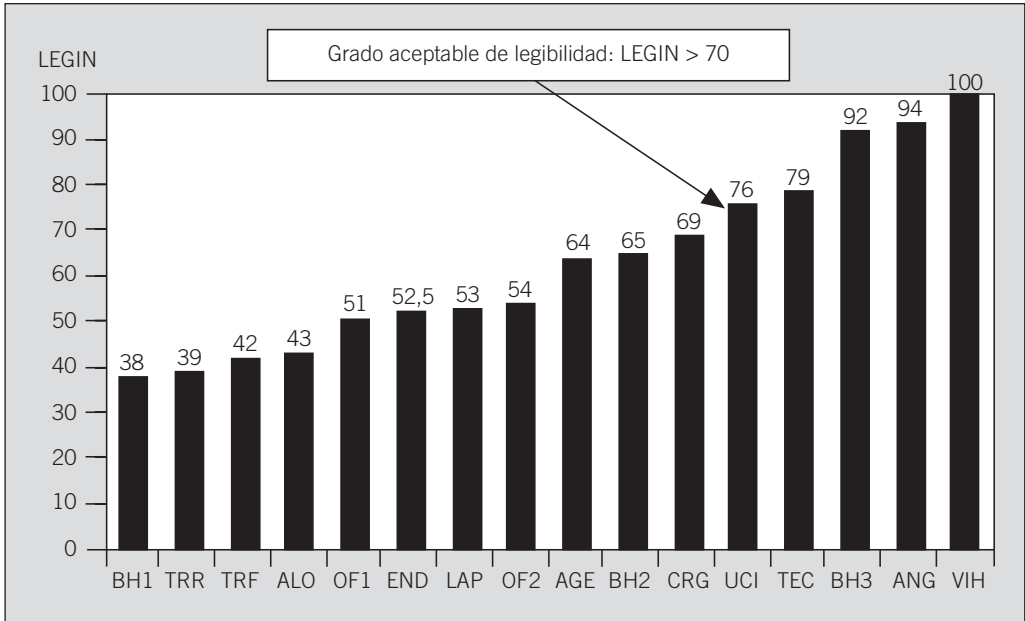


Fig. 3. Representación de los valores del índice de Flesch y del índice de complejidad oracional obtenidos al analizar los formularios escritos de consentimiento informado. AGE: anestesia general; ALO: anestesia locorregional; ANG: angiografía; BH1: biopsia hepática 1; BH2: biopsia hepática 2; BH3: biopsia hepática 3; CRG: cirugía (formato semiespecífico); END: endoscopia; LAP: laparoscopia; OF1: cirugía de cataratas; OF2: angiofluoresceína; TEC: electroshock; TRF: transfusión; TRR: trasplante renal; UCI: unidad de cuidados intensivos; VIH: test del VIH.

para todos<sup>31</sup>. Además el consentimiento informado es un proceso relativamente nuevo para la ética de los sanitarios, tradicionalmente aferrados a concepciones paternalistas de la relación médico-paciente. Por todo ello, no es extraño que los formularios que actualmente se emplean sean tan poco legibles. Pero el grado de esfuerzo que se realice por ir abandonando progresivamente la jerga críptica y las actitudes patrimonialistas hacia los pacientes sí es un buen indicador del calibre de la ética profesional del sanitario, sobre todo de su nivel de respeto por la autonomía de los enfermos. Por eso, lo que sí sería éticamente cuestionable es que esos formularios no se modificasen para hacerlos más comprensibles, o que el análisis de los formularios usados en nuestro país en un plazo de uno a 2 años revelara datos tan pobres como los obtenidos en este estudio.

Resulta imprescindible señalar que el análisis de la calidad de los formularios escritos de consentimiento informado no se agota en el análisis de la legibilidad. Hay dos variables más que deben tenerse en cuenta. Una es la cantidad de información y su contenido, otra es la dificultad ter-

minológica. La primera debe ser controlada por el personal sanitario que realiza las técnicas: son ellos quienes mejor saben, aplicando el *criterio de la persona razonable*<sup>1</sup>, de qué hay que informar al paciente. No es una tarea del comité asistencial de ética tomar decisiones al respecto. La segunda precisaría sistemas de medición más objetivos pero, visto lo sucedido con el *índice de complejidad del vocabulario*, probablemente son difíciles de lograr. Posiblemente decidir si un determinado término debe ser considerado o no un *tecnicismo* innecesario en un determinado contexto es algo en buena medida consensuable<sup>6</sup>, tarea en la que el comité asistencial de ética sí puede aportar sugerencias relevantes.

Por último, debemos señalar los evidentes límites de este estudio. El más importante estriba en que los instrumentos de análisis utilizados dependen de un programa informático concreto, lo que limita su uso a los que puedan disponer de él. Por otra parte, el programa sólo tiene, por ahora, versión en castellano. En el caso de que la casa comercial realizara versiones en las otras lenguas oficiales en el Estado español, cosa du-

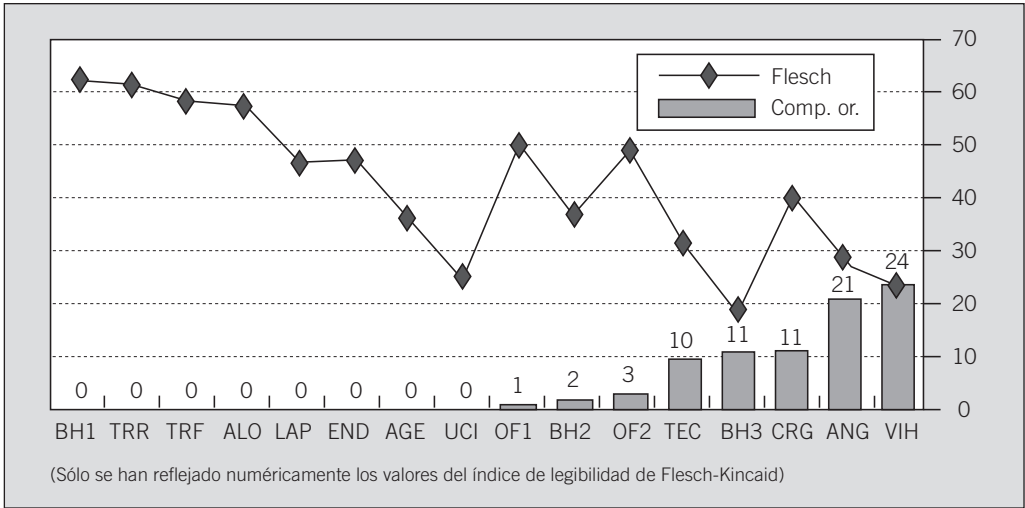


Fig. 4. Representación gráfica de los valores del LEGIN en el análisis de los formularios escritos de consentimiento informado. AGE: anestesia general; ALO: anestesia locorregional; ANG: angiografía; BH1: biopsia hepática 1; BH2: biopsia hepática 2; BH3: biopsia hepática 3; Comp. or.: puntuación del índice de complejidad de oraciones; CRG: cirugía (formato semiespecífico); END: endoscopia; LAP: laparoscopia; OF1: cirugía de cataratas; OF2: angiofluoresceína; TEC: electroshock; TRF: transfusión; TRR: trasplante renal; UCI: unidad de cuidados intensivos; VIH: test del VIH.

dosa, habría que revisar la validación de los índices de legibilidad. Por último, el limitado número de muestras estudiadas hace que los resultados y las recomendaciones que de ellos se derivan, sólo tengan un valor orientativo general. No son normas fijas, ni garantizan inequívocamente que el nivel de legibilidad sea el adecuado para cada caso. Se precisan, pues, nuevos estudios para poder evaluar el grado de sensibilidad y especificidad de los índices empleados y para revisar la operatividad de las recomendaciones realizadas sobre los valores mínimos aceptables de cada uno de ellos.

## BIBLIOGRAFÍA

1. Simón Lorda P, Concheiro Carro L. El consentimiento informado: teoría y práctica (I). *Med Clin (Barc)* 1993; 100: 659-663.
2. Grundner TM. Two formulas for determining the readability of subject consent forms. *Am Psychol* 1978; 33: 773-775.
3. Levine RJ. *Ethics and regulation of clinical research* (1.ª ed.). Baltimore - Munich: Urban & Schwarzenberg, 1986; 138-139.
4. Powers RD. Emergency department patient literacy and the readability of patient-directed materials. *Ann Emerg Med* 1988; 17: 124-126.
5. Albert T, Chadwick S. How readable are practice leaflets? *Br Med J* 1992; 305: 1.266-1.268.
6. Simón Lorda P, Concheiro Carro L. El consentimiento informado: teoría y práctica (y II). *Med Clin (Barc)* 1993; 101: 174-182.
7. Word Perfect. Versión 6.1. para Windows. Orem-Utah: Novell Inc., 1994.
8. Cervantes Saavedra M de. *El ingenioso hidalgo Don Quijote de la Mancha*. Barcelona: Círculo de Lectores, 1969; 41-42.
9. Platón. *Diálogos IV*. La República. Edición de Conrado Eggers Lan. Madrid: Gredos, 1986; 57-59.
10. Nueva Biblia Española. Edición de Luis A. Schökel y Juan Mateos (2.ª ed.). Madrid: Cristiandad, 1977; 1,1-2,4.
11. Terán R de. *Los mifenses* (3.ª ed.). Col. El Barco de Vapor. Madrid: SM, 1985; 7-10.
12. *Medicina Clínica* 1995; 104 (6).
13. *Cirugía Española* 1995; 57 (6).
14. *Lancet* (ed. esp.) 1994; 24 (5).
15. *Jano. Medicina y Humanidades* 1994; XLVII (1.100).
16. *Revista Semana*. 7 al 13 de agosto de 1995.
17. *Muy Interesante* 1995; (171).
18. *Revista Pronto*; 7 al 13 de agosto de 1995.
19. *El País Semanal* 1995; Año XIX. Tercera época (198).
20. Alcalá-Zamora J, Abizanda R, Moreno JL, Miravet V. Obtención del consentimiento informado en medicina intensiva. *JANO* 1995; XLVIII (1.117): 73-79.
21. Simón D, Simón P, Rodríguez A, Concheiro L. El consentimiento informado en psiquiatría: un formulario escrito de consentimiento para terapia electroconvulsiva (TEC) o electroshock. *An Psiquiatría* 1992; 8: 245-256.

22. Simón Lorda P, Barrio Cantalejo IM. Bioética, información e infección por el VIH. FMC - Formación Médica Continuada en Atención Primaria 1994; 1: 299-309.
23. Word Perfect v. 6.1. Guía para el usuario. Orem-Utah: Novell Inc., 1994; 367-369.
24. Grundner TM. On the readability of surgical consent forms. N Engl J Med 1980; 302: 900-902.
25. Morrow GR. How readable are subject consent forms? JAMA 1980; 244: 56-58.
26. Riecken HW, Ravich R. Informed consent to biomedical research in Veterans Administration hospitals. JAMA 1982; 248: 344-348.
27. Baker MT, Taub HA. Readability of informed consent forms for research in a Veterans Administration medical center. JAMA 1983; 250: 2.646-2.648.
28. Tarnowski KJ, Allen DM, Mayhall C, Kelly PA. Readability of pediatric biomedical research informed consent forms. Pediatrics 1990; 85: 58-62.
29. Ogloff JRP, Otto RK. Are research participants truly informed? Readability of informed consent form used in research. Ethics & Behavior 1991; 1: 239-252.
30. Priestley KA, Campbell C, Valentine CB, Denison DM, Buller NP. Are patient consent forms for research protocols easy to read? Br Med J 1992; 305: 1.263-1.264.
31. Smith T. Information for patients: Writing simple English is difficult, even for doctors. Br Med J 1992; 305: 1.242.

---

# Análisis de las hojas de información al paciente para la obtención de su consentimiento informado en ensayos clínicos

---

Juan Pablo Ordovás Baines<sup>a</sup>, Eduardo López Briz<sup>b</sup>, Elena Urbietta Sanz<sup>c</sup>, Rafael Torregrosa Sánchez<sup>d</sup> y N. Víctor Jiménez Torres<sup>a,e</sup>

<sup>a</sup>Comité Ético de Investigación Clínica. Servicio de Farmacia. Hospital Universitario Dr. Peset. Valencia. <sup>b</sup>Servicio de Farmacia. Hospital Universitario La Fe. Valencia. <sup>c</sup>Servicio de Farmacia. <sup>d</sup>Unidad de Garantía de Calidad. Hospital General Universitario. Valencia. <sup>e</sup>Departamento de Farmacia y Tecnología Farmacéutica. Universidad de Valencia.

España ha experimentado en las dos últimas décadas un importante incremento en el número y calidad de los ensayos clínicos con medicamentos y, de forma paralela, un desarrollo avanzado de la legislación al respecto<sup>1-4</sup>. Actualmente, disponemos de una de las normativas legales más avanzadas sobre investigación clínica con medicamentos. Desgraciadamente, su existencia, aunque imprescindible, no es suficiente garantía de una conducta moral, como la historia se encarga de recordarnos<sup>5-9</sup>.

La regulación ética y legal del consentimiento informado en los ensayos clínicos presenta una indudable importancia. El derecho a la autonomía del paciente se considera en nuestra sociedad un derecho fundamental, y por ello la reciente legislación española recoge de forma bastante detallada los requerimientos mínimos para su aseguramiento<sup>4</sup>. Resulta imprescindible que el sujeto otorgue libremente su consentimiento informado antes de que pueda ser incluido en un ensayo clínico.

La comprensión de las palabras en un contexto determinado es un complejo proceso que incluye la lógica, el lenguaje y la experiencia. La capacidad de entendimiento de la información escrita que recibe un paciente puede mejorarse significativamente ajustando la legibilidad formal del texto a su nivel de lectura<sup>10</sup>. Por ello, para que el consentimiento sea considerado válido, el paciente debe recibir la cantidad y calidad de información adecuada a su grado de comprensión<sup>11</sup>. Sin embargo, pese a que en los países de habla inglesa es frecuente el análisis de legibilidad de los materiales escritos que se proporcionan a los pacientes y los familiares, y en especial los relacionados con temas terapéuticos y de investigación clínica<sup>10,12-15</sup> utilizando una gran variedad de técnicas<sup>16-18</sup>, en nuestro país y en nuestro idioma

no se ha efectuado hasta la fecha un análisis detallado de la legibilidad de la información escrita que se proporciona a los pacientes reclutados para un ensayo clínico. Esto es tanto más preocupante cuando en la gran mayoría de estudios anglosajones los materiales escritos presentan una muy pobre legibilidad formal, esto es, una notable complejidad.

Los objetivos del presente trabajo son, por una parte, establecer el grado de legibilidad formal de la hoja de información que el paciente recibe cuando se solicita su participación en un ensayo clínico y, por otra, analizar la cantidad y calidad de la información que se aporta en dicha hoja, de acuerdo con los requerimientos estándares establecidos en España<sup>4</sup>.

## Material y métodos

---

El estudio se llevó a cabo en dos hospitales públicos, generales y universitarios de la Comunidad Valenciana: uno de 700 camas dependiente de la Diputación Provincial y otro de 550 camas perteneciente a la Conselleria de Sanitat. En ambos centros existe un comité ético de investigación clínica (CEIC) debidamente acreditado, que se responsabiliza de la evaluación de los protocolos de ensayo clínico a realizar en cada centro y su área de salud, de acuerdo con la legislación vigente.

Se incluyeron en el estudio la totalidad de los protocolos de ensayos clínicos aprobados por los CEIC de ambos hospitales entre enero de 1994 y diciembre de 1996 (n = 101), en los que figuraba, dado que es condición *sine qua non* para su aprobación, una hoja de información al paciente (HIP). Cuatro de los autores (OB, LB, US y TS), miembros de los CEIC de uno u otro hospital, se encargaron de la recogida de datos y su procesamiento estadístico. Tras verificar el ajuste de las variables a la normalidad mediante la prueba de Kolmogorov-Smirnov, se utilizaron

TABLA I  
CLASIFICACIÓN DEL GRADO DE COMPLEJIDAD DE LAS HOJAS  
DE INFORMACIÓN AL PACIENTE (HIP)

Grado	Nivel de complejidad del texto	Índice de Flesch	Nivel de estudios necesarios para su comprensión
E	Elevado	≤ 10	Superiores
D	Medio	0 a -10	Medios-superiores*
C	Comprensible	0-10	Medios
B	Sencillo	10-40	Secundarios
A	Muy sencillo	> 40	Primarios

\*Se agrupan en este nivel los estudios de grado medio relacionados con la sanidad. Modificada de Simón<sup>11</sup>.

los siguientes estadísticos inferenciales: la t de Student de dos colas para muestras independientes en la comparación de dos medias (previa prueba de Levene para comparar la igualdad de variancias), el análisis de la variancia de una vía para muestras independientes (con la corrección de Bonferroni) en la comparación de más de dos medias y la  $\chi^2$  para variables cualitativas o categóricas. Para el análisis estadístico se empleó el programa informático SPSS® para Windows (SPSS Inc., versión 6.0.1).

#### Características de los ensayos clínicos

En cada ensayo clínico se recogieron las siguientes características: servicio clínico al que pertenecía el investigador principal o donde se iba a llevar a cabo el ensayo, fase del ensayo (I-IV), diseño del enmascaramiento (abierto, simple ciego y doble ciego) y control del ensayo (con placebo, con otro fármaco y no controlado).

#### Características de la HIP

Las HIP se analizaron en cuanto a la calidad de la información, la legibilidad formal y la complejidad técnica del vocabulario empleado:

**Calidad de la información.** Se comprobó si las HIP contenían o no información adecuada y suficiente en relación con los aspectos mínimos (12 ítems) reflejados en la legislación española<sup>4</sup>. Para ello, fueron evaluados por duplicado por dos personas diferentes, y se efectuó un consenso con un tercer autor en caso de discrepancia. Se controló también si las hojas de información eran firmadas por el paciente o si, por el contrario, el paciente debía firmar únicamente el impreso de consentimiento informado de forma separada. También se determinó si, como es preceptivo en la legislación española, la hoja de consentimiento informado por escrito indicaba que se había

leído la HIP o si se había omitido esta observación. Se recogió, finalmente, el porcentaje de HIP que aportaban explícitamente el número total de pacientes que iban a participar en el estudio.

**Legibilidad formal.** Para la cuantificación de la legibilidad formal, se utilizó el índice de legibilidad descrito por Flesch<sup>16</sup> calculado manualmente y por duplicado. El índice obtenido mediante este método permite cuantificar la dificultad de lectura de un texto en función de la estructura lingüística del mismo (largura de palabras y frases). Aunque originalmente descrito para la lengua inglesa con valores de 0 a 100 (a mayor índice, mayor sencillez o legibilidad del texto), su aplicación al castellano por Simón<sup>11</sup> proporcionó valores entre -42 y 53, correspondiéndose igualmente a mayor número una mayor legibilidad.

Cada HIP se analizó para la obtención de las siguientes variables: número de páginas de que constaba, índice de Flesch, largura media de palabras, largura media de frases y número total de palabras.

Para relacionar la legibilidad (calculada como índice de Flesch) con el nivel de complejidad de las HIP se confeccionó una escala, basada en los valores proporcionados por Simón<sup>11</sup>, que se recoge en la tabla I. A efectos comparativos, se compararon los resultados de legibilidad obtenidos en la muestra de ensayos clínicos estudiados con los obtenidos de diversos textos de dificultad intrínseca variable, publicados previamente en esta misma revista<sup>11</sup>.

**Complejidad técnica del vocabulario.** Para el análisis de la complejidad técnica se solicitó la colaboración de dos voluntarios, licenciados universitarios y ajenos a la profesión sanitaria, de 37 y 39 años respectivamente, que identificaron de forma simple aquellas palabras de la totalidad de las HIP cuyo significado no eran capaces de explicar. Para evitar la distorsión de los resulta-

dos, no se tuvieron en cuenta los nombres de medicamentos ni los términos que se repetían a lo largo de la HIP.

## Resultados

### Características de los ensayos clínicos

La distribución de los ensayos clínicos por servicios clínicos se recoge en la tabla II.

La distribución por fases demostró una mayoría de ensayos clínicos de fase III y IV (el 62,4 y el 29,7 %, respectivamente), correspondiendo a las fases I y II porcentajes de menor entidad (el 2,0 y el 5,9 %, respectivamente). Por su parte, el enmascaramiento se repartió a partes prácticamente iguales entre los ensayos clínicos doble ciegos (50,5 %) y los abiertos (49,5 %). Respecto al tipo de control, 53 ensayos (52,5 %) utilizaban un fármaco activo para el grupo control, 19 ensayos (18,8 %) emplearon placebo y 29 (28,7 %) fueron no comparativos.

### Características de la HIP

*Calidad de la información.* Los ítems que más negativamente se puntuaron fueron el modo de contacto con el investigador principal, el balance riesgo-beneficio para el paciente, la identificación del investigador principal, la descripción de tratamientos alternativos, el modo de compensación económica y la confidencialidad de los datos del paciente, como se expone en la tabla III.

TABLA II  
DISTRIBUCIÓN POR SERVICIOS DE  
LOS ENSAYOS CLÍNICOS ESTUDIADOS

Servicio	Ensayos clínicos estudiados (%) (n = 101)
Alergia	2 (2)
Anestesiología	1 (1)
Atención primaria	1 (1)
Cardiología	8 (7,9)
Cirugía vascular	1 (1)
Cuidados intensivos	2 (2)
Dermatología	12 (11,9)
Digestivo	8 (7,9)
Endocrinología	4 (4)
Estomatología	2 (2)
Hematología	9 (8,9)
Medicina interna	7 (6,9)
Nefrología	1 (1)
Neumología	3 (3)
Neurocirugía	1 (1)
Neurología	15 (14,9)
Oncología	9 (8,9)
Otorrinolaringología	4 (4)
Pediatría	2 (2)
Radiodiagnóstico	5 (5)
Reumatología	2 (2)
Urología	2 (2)

El análisis de la calidad de la información, estratificada por fases, en aquellos ítems donde hubo una respuesta negativa superior al 15 %,

TABLA III  
ANÁLISIS DEL CONTENIDO DE LAS HOJAS DE INFORMACIÓN AL PACIENTE (HIP) (N = 101)

	Cumple/no cumple	Porcentaje de incumplimiento
1. Se describen los objetivos	95/6	5,9
2. Se describe la metodología	95/6	5,9
3. Se describe el tratamiento que puede ser utilizado	96/5	5,0
4. Se indican riesgos, molestias y efectos adversos	92/9	8,9
5. Se indica el balance riesgo-beneficio	50/51	50,5
6. Se indican los tratamientos alternativos posibles	62/39	38,6
7. Se precisa la voluntariedad de la participación	93/8	7,9
8. Se especifica la retirada voluntaria	95/6	5,9
9. Se especifica que la retirada no afectará a la calidad del tratamiento	87/14	13,9
10. Se indica qué personas tendrán acceso a la información y la confidencialidad	80/21	20,8
11. Se describe el modo de compensación económica en caso de daño o lesión	68/33	32,3
12. Se indica el médico responsable del ensayo	60/51	40,6
13. Se especifica cómo contactar con el responsable en caso de urgencia	49/52	51,5

TALBA IV  
PARÁMETROS GLOBALES DE LEGIBILIDAD FORMAL DE LAS 101 HOJAS DE INFORMACIÓN AL PACIENTE (HIP) ESTUDIADAS

	<i>Media (DE)</i>	<i>Mínimo</i>	<i>Máximo</i>
Legibilidad formal (índice de Flesch)	-12,72 (11,85)	-48,00	38,76
Número de páginas	2,60 (1,25)	1,00	6,00
Longitud de frases	25,60 (5,61)	15,75	49,20
Longitud de párrafos	228,45 (12,82)	171,50	257,30
Número total de palabras	812,75 (459,89)	132,00	3.050,00

DE: desviación estándar.

tras excluir las fases I y II por lo escaso de la muestra, no arrojó diferencias significativas (prueba de la  $\chi^2$ ) entre las fases III y IV en lo que respecta a la presencia o no en las HIP de adecuado análisis riesgo-beneficio, alternativas de tratamiento, posibilidad de compensación económica en caso de lesiones, identificación del responsable del ensayo clínico y modo de contactar con él. Por el contrario, se obtuvo una diferencia estadísticamente significativa ( $p < 0,01$ ) en la mención de la confidencialidad, que fue sensiblemente mayor en los ensayos clínicos en fase III (67,7 %) en relación con los ensayos de fase IV (32,3 %).

En un porcentaje elevado de las HIP (20,8 %), no se mencionó que el paciente había leído la información que se le aportaba, y únicamente 38 (37,6 %) de las HIP recogieron el número de pacientes que iban a participar en el estudio.

*Legibilidad formal.* Los parámetros de legibilidad formal globales pueden observarse en la tabla IV. En la tabla V, se desglosan por fases y se refleja también el número total de palabras de las HIP analizadas. En la figura 1, se comparan los resultados de legibilidad obtenidos, a nivel global y por fases, con los de los textos analizados previamente por Simón<sup>11</sup>.

TABLA V  
ALGUNOS PARÁMETROS  
DE LEGIBILIDAD EN FUNCIÓN DE  
LA FASE DEL ENSAYO CLÍNICO

<i>Fase (n)</i>	<i>Tm (DE)</i>	<i>IF (DE)</i>
I (2)	1.358 (398,81)	-23,65 (12,03)
II (6)	617,17 (221,68)	-21,52 (11,99)
III (63)	868,86 (523,49)	-11,36 (12,76)
IV (30)	697,70 (283,76)	-13,07 (8,83)

DE: desviación estándar; IF: índice de Flesch medio; n: número de ensayos; Tm: número total de palabras medio;

Aunque sin significación estadística ( $p = 0,12$ ), se observa una cierta tendencia a mejorar la legibilidad formal al progresar en la fase de desarrollo clínico del medicamento (mejor legibilidad en las fases más avanzadas de investigación).

El nivel de complejidad en la lectura de las HIP, agrupado según los grados descritos en la tabla I, y expresado como el nivel estimado de estudios necesario para su comprensión, se recoge en la tabla VI. Según este análisis, un 97 % de las HIP requieren para su comprensión niveles de estudios medios o superiores.

Los servicios clínicos con mayor número de ensayos en los que fue particularmente elevada la proporción de necesidad de estudios superiores para la comprensión de las HIP fueron: medicina digestiva (87,5 %), hematología (77,8 %), neurología (86,7 %) y oncología (66,7 %).

Respecto al enmascaramiento, no se han obtenido diferencias estadísticamente significativas entre el índice de Flesch calculado para los ensayos clínicos no ciegos (-14,57;  $n = 50$ ) frente a los ciegos (-10,90;  $n = 51$ ) ( $p = 0,12$ ; intervalo de confianza del 95 % = 8,33/0,97). Algo parecido ocurre cuando se estudia la legibilidad de los ensayos en función del tipo de procedimiento de control empleado, es decir, no existen diferencias ( $p = 0,089$ ) entre los tres grupos analizados: con-

TABLA VI  
DISTRIBUCIÓN DE LAS HOJAS DE  
INFORMACIÓN AL PACIENTE (HIP)  
SEGÚN EL NIVEL DE ESTUDIOS  
NECESARIO PARA SU COMPRENSIÓN

<i>Nivel de estudios</i>	<i>Número (%) (n = 101)</i>
Primarios	0 (0)
Secundarios	3 (3)
Medios	7 (6,9)
Medios-superiores	25 (24,8)
Superiores	66 (65,3)



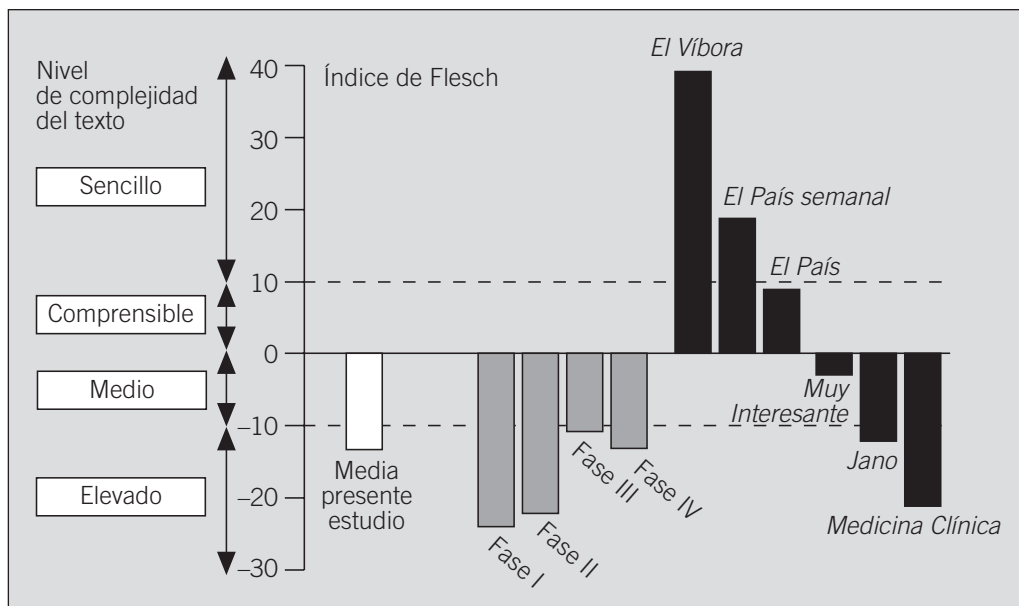


Fig. 1. Resultados medios obtenidos de legibilidad (índice de Flesch) global y por fases comparados con los resultados de textos de diversa dificultad. Tomada de Simón<sup>11</sup>.

trol con fármaco activo (-12,72; n = 51), control con placebo (-8,08; n = 19) y ensayo no controlado (-15,75; n = 29).

**Complejidad técnica del vocabulario.** El número de palabras cuyo significado desconocían los dos voluntarios que realizaron la prueba fue de 3,3 (desviación estándar [DE] 2,3), por cada 1.000 palabras de texto (mínimo 0 y máximo 12), esto es, un 0,3%.

## Discusión

La comprensión de la HIP va a depender principalmente de cinco variables: a) nivel sociocultural del paciente; b) nivel de sus conocimientos sanitarios; c) grado de motivación del paciente; d) nivel de complejidad formal o estructural (legibilidad) del texto, y e) nivel de complejidad del vocabulario empleado en el texto.

En principio, los problemas de comprensión de las HIP no parecen deberse a una elevada complejidad de vocabulario, si bien es preciso recordar que este análisis se ha efectuado únicamente con dos voluntarios con estudios superiores.

El índice de Flesch medio obtenido (-12,72; DE 11,85) es preocupantemente bajo. Simón<sup>11</sup>, al aplicar este método por primera vez para eva-

luar formularios de consentimiento en castellano, concluía que el nivel óptimo de legibilidad debía situar al índice de Flesch por encima de 10. Sin embargo, en un reciente artículo, el mismo autor<sup>18</sup> considera mejor afirmar que la puntuación del índice de Flesch debe ser siempre superior a 10, puesto que es por encima de este valor donde se sitúan los textos y las publicaciones destinadas al ciudadano español medio.

¿Qué puede interpretarse de un índice de Flesch medio de -13 encontrado en una muestra de ensayos clínicos desarrollados en nuestro entorno? (fig. 1). La muestra de ensayos analizados parece lo suficientemente amplia como para avalar la fiabilidad de los resultados globales, que soportan la idea de que el material escrito analizado presenta un nivel inaceptable de dificultad de lectura. Es inevitable plantear ante estos datos un potencial problema ético relevante: los participantes que firmaron el consentimiento informado han podido no ser adecuadamente informados sobre los ensayos clínicos en que participaron. Si ello fuera así, su consentimiento informado debería considerarse inválido. Sin embargo, es difícil evaluar el grado en el que los pacientes entendieron la HIP antes de otorgar su consentimiento. Para ello, sería necesario efectuar un estudio prospectivo mediante una entre-

vista a los pacientes posterior a la lectura de la HIP. Tal estudio, no desarrollado por el momento, determinaría la relación entre el índice de Flesch y la comprensión real del paciente, es decir, entre legibilidad y comprensión.

La metodología de la determinación del grado de legibilidad formal mediante el índice simple de Flesch no es de excesiva dificultad, y es relativamente firme desde el punto de vista de la consistencia interobservador. La utilización de aplicaciones computarizadas<sup>18</sup> exige un no desdoblado número de normas prácticas para la correcta obtención del índice de Flesch y de los valores de complejidad lingüística<sup>18</sup>, y su reproducibilidad y variabilidad interobservador parecen más que discutibles<sup>19</sup>.

La nueva regulación española deposita en los CEIC una gran responsabilidad en la evaluación de los protocolos de ensayos clínicos, especialmente desde el punto de vista ético<sup>4</sup>. El hecho de que todos los ensayos clínicos analizados en el presente estudio hubieran sido aprobados previamente por un CEIC debe hacernos reflexionar acerca de cómo se pueden dar por válidas HIP con índices de Flesch de -48, es decir, de una complejidad similar a una revista biomédica especializada y un 500 % superior al estándar propuesto. De hecho, en el trabajo cotidiano de un CEIC, uno de los problemas más frecuentes que surgen al evaluar los protocolos de los ensayos clínicos lo constituyen las HIP incompletas o pobremente redactadas<sup>20</sup>. Si a esto sumamos que, con el tiempo y con la repetida exposición a la terminología propia de la investigación clínica, los miembros del CEIC (incluso los ajenos a la profesión sanitaria) tienden a considerar la redacción de las HIP más legible<sup>15</sup>, los resultados son más desalentadores si cabe.

Las recomendaciones de Ogloff y Otto<sup>12</sup> para los investigadores podrían haber mejorado, si se hubieran aplicado, la legibilidad y la comprensión de las HIP: *a)* intentar hacerlas lo más legibles posible; *b)* hacer leer la HIP a alguien de la misma educación y edad que aquel al que va dirigido el ensayo para determinar si puede entenderla; *c)* solicitar a los potenciales participantes que expliquen los procedimientos del ensayo con sus propias palabras, y *d)* proporcionar a los pacientes un tiempo mínimo de 15 min entre el momento en que la HIP es leída (y explicada) y el momento en que se pregunta si está de acuerdo en participar. Aún más, se ha demostrado que los pacientes oncológicos incrementan su grado de comprensión sobre el tratamiento que se les propone cuando el formulario de consentimiento informado se les proporciona con la antelación su-

ficiente para que lo puedan leer en su casa, antes de firmarlo en el hospital<sup>21</sup>. Las citadas recomendaciones, y otras similares, deben dirigirse prioritariamente a los promotores e investigadores de los ensayos, dado que se ha demostrado que los CEIC tienen una capacidad muy limitada para mejorar la legibilidad de las HIP<sup>14,22</sup>. Representantes de la Food and Drug Administration (FDA)<sup>23</sup> han expresado recientemente que las deficiencias con más frecuencia encontradas en los formularios para el consentimiento informado de los ensayos clínicos en EE.UU. son, por este orden: *a)* no identificación de la persona de contacto en caso de urgencia o para resolver cuestiones sobre la investigación o los derechos del paciente; *b)* inadecuado establecimiento de la confidencialidad; *c)* incompleta descripción de los procedimientos de investigación, y *d)* incompleta descripción de las alternativas terapéuticas.

Los resultados del presente estudio, en relación con la cantidad y calidad de información aportada a los pacientes, coinciden con lo expresado por los representantes de la FDA en el primer punto: la identificación o el modo de contacto con el investigador o responsable de atender al paciente tras su inclusión. La HIP debe entregarse personalmente al paciente, que debería disponer de tiempo suficiente para tomar una decisión<sup>24</sup>. Sería recomendable que la HIP no sólo quedara en poder del paciente, sino que fuera firmada conjuntamente por él mismo y por el investigador, lo que no se da más que en dos de los protocolos estudiados. El hecho de que más del 20 % de los impresos de consentimiento informado analizados no reflejen que el paciente ha leído la HIP que preceptivamente figura en el protocolo del ensayo, debe ser también motivo de reflexión por las posibles repercusiones legales que de ello pudieran derivarse dado que la legislación exige explícitamente tal indicación<sup>4</sup>.

Comunicar al paciente el número de participantes previsto en el ensayo es una recomendación que los CEIC (Institutional Review Boards) en EE.UU. efectúan con frecuencia a sus investigadores<sup>25</sup>. Esto tiene su lógica, dado que de alguna manera se incrementa la implicación del individuo con el proyecto y se le aproxima a la realidad de la investigación. El bajo número de HIP que aportan este dato puede dar una idea del poco rigor con el que se elaboran.

Un déficit de información importante es el encontrado en el apartado de balance riesgo-beneficio que el ensayo representa para el paciente. Si bien es un parámetro relativamente subjetivo de valorar, en más del 50 % de las HIP no se explicaban de forma mínimamente aceptable los ries-

gos y los beneficios para el paciente derivados del estudio. Así mismo, en muchas ocasiones el apartado de tratamientos alternativos (en el que se ha encontrado un incumplimiento cercano al 40 %) se confunde con descripción del tratamiento del grupo en estudio y del grupo control del ensayo, cuando en realidad en la HIP se requiere explicar al paciente los tratamientos alternativos fuera del ensayo, que se encuentran disponibles en ese momento para tratar su proceso patológico que se pretende estudiar. Por otra parte, y coincidente con lo referido en EE.UU.<sup>23</sup>, la confidencialidad de los datos del paciente y los procedimientos para mantenerla a lo largo de la investigación no quedan adecuadamente descritos en más del 20 % de las HIP.

Finalmente, debe señalarse como de suma trascendencia que el 90,1 % de las HIP necesitan para su comprensión que el paciente haya cursado estudios medios o superiores (tabla VI). Frente a este sorprendente dato, contrasta el hecho de que, en 1991, sólo un 34,2 % de la población española poseía título académico de segundo grado y más del 3 % eran considerados analfabetos<sup>26</sup>. Sería de gran ayuda que los promotores tomaran conciencia del hecho y se esforzaran en redactar las HIP según las recomendaciones recogidas en la bibliografía al respecto<sup>27</sup>, analizando, previamente a su inclusión en el protocolo y mediante medios manuales o informatizados, la legibilidad formal de las mismas.

La práctica del consentimiento informado del paciente, ya en investigación clínica, ya en la práctica médica habitual, no ha sido suficientemente analizada en nuestro país<sup>28</sup>. Pese a los muchos mitos éticos existentes en el ámbito de los ensayos clínicos<sup>29</sup>, se debe reconocer que las hojas de información y los formularios de consentimiento del paciente tienen un importante valor; le brindan al paciente la oportunidad de leer la información, en el momento de su decisión y con posterioridad a la misma. Vale la pena esforzarse en contestar a la pregunta de por qué participan los pacientes en un ensayo clínico. Mattson et al<sup>30</sup> han encontrado que más del 40 % de los pacientes entrevistados citaba como beneficio más evidente de su participación en un ensayo la adquisición de información clínica adicional sobre su enfermedad, y un porcentaje menor citaba una mejor educación sobre su enfermedad y su control. A esto debe sumarse el alto porcentaje de pacientes que declaraba participar en los ensayos clínicos para ayudar a otros, especialmente a las personas que padecen su misma enfermedad<sup>30</sup>. De este modo, podría argumentarse que un método eficaz para fomentar genérica-

mente la participación de los pacientes en los ensayos clínicos (realmente baja en la actualidad) es informarles adecuadamente de los objetivos del ensayo, de las alternativas de tratamiento, de los tratamientos propuestos en el ensayo, del control de la enfermedad y del curso del tratamiento terapéutico, e incluso, como se ha sugerido, de los resultados de los ensayos<sup>31</sup>. Por el contrario, resulta contraproducente que el paciente no comprenda partes importantes de las razones, características y alternativas del ensayo, dado que sería motivo de rechazo o de baja participación. Para este fin, la HIP puede ser un importante instrumento en manos del investigador.

## BIBLIOGRAFÍA

1. Real Decreto 944/1978, de 14 de abril, sobre regulación de los ensayos clínicos de productos farmacéuticos y preparados medicinales. Ministerio de Sanidad y Seguridad Social. BOE n.º 108, 1978.
2. Orden Ministerial 29605/1982, de 3 de agosto de 1982, que desarrolla el Real Decreto 944/ 1978 sobre ensayos clínicos en humanos. Ministerio de Sanidad y Consumo. BOE n.º 192, 1982.
3. Ley 14/1986, de 25 de abril, del medicamento. Ministerio de Sanidad y Consumo. BOE n.º 78, 1986.
4. Real decreto 561/1993, de 16 de abril, por el que se establecen los requisitos para la realización de ensayos clínicos con medicamentos. Ministerio de Sanidad y Consumo. BOE n.º 114, 1993.
5. Sass HM. The Nüremberg code, German law, and prominent physician-thinkers. *JAMA* 1997; 277: 709.
6. Lizón J. Ética y ensayo clínico en oncología. En: Díaz Rubio E, editor. El ensayo clínico en oncología. Madrid: Ergón S.A., 1995; 53-67.
7. Enserink M. Fraud and ethics charges hit stroke drug trial. *Science* 1996; 274: 2.004-2.005.
8. Katz J. The Nüremberg code and the Nüremberg trial. A reappraisal. *JAMA* 1996; 276: 1.662-1.666.
9. The Nüremberg code. *JAMA* 1996; 276: 1.691.
10. Eaton ML, Holloway RL. Patient comprehension of written drug information. *Am J Hosp Pharm* 1980; 37: 240-243.
11. Simón P. El consentimiento informado: teoría y práctica (y II). *Med Clin (Barc)* 1993; 101: 174-182.
12. Ogloff JPR, Otto RK. Are researchs participants truly informed? Readability of informed consent forms used in research. *Ethics Behavior* 1991; 1: 239-252.
13. Baker MT, Taub HA. Readability of informed consent forms for research in a veterans administration medical center. *JAMA* 1983; 250: 2.646-2.648.
14. Murgatroyd RJ, Cooper RM. Readability of informed consent forms. *Am J Hosp Pharm* 1991; 48: 2.651-2.652.

15. Tarnowski KJ, Allen DM, Mayhall C, Kelly PA. Readability of pediatric biomedical research informed consent forms. *Pediatrics* 1990; 85: 58-62.
16. Grundner TM. Two formulas for determining the readability of subject consent forms. *Am Psychologist* 1978; 33: 773-775.
17. Bradley B, Singleton M, Li Wan Po A. Readability of patient information leaflets on over-the-counter medicines. *J Clin Pharm Ther* 1994; 19: 7-15.
18. Simón P, Barrio MI, Concheiro L. Legibilidad de los formularios escritos de consentimiento informado. *Med Clin (Barc)* 1997; 107: 524-529.
19. Mailloux SL, Johnson ME, Fisher DG, Pettibone TJ. How reliable is computerized assessment of readability? *Comput Nurs* 1995; 13: 221-225.
20. University of Illinois at Urbana-Champaign Institutional Review Board. Handbook for investigators: for the protection of human subjects in research. Urbana-Champaign: Ed. University of Illinois, 1995.
21. Morrow G, Gootnick J, Schmale A. A simple technique for increasing patient's knowledge of informed consent to treatment. *Cancer* 1978; 42: 793-799.
22. Grossman SA, Piantadosi S, Covahey C. Are informed consent forms that describe clinical oncology research protocols readable by most patients and their families? *J Clin Oncol* 1994; 12: 2.211-2.215.
23. Briley J. Informed consent pitfalls from FDA and Hopkins IRB chairman. *Clin Trials Advisor* 1996; 1: 1-2.
24. Vallvé C. La buena práctica clínica y la caja de Pandora (II). El consentimiento informado. *Med Clin (Barc)* 1997; 108: 19-22.
25. Virginia Polytechnic Institute and State University. Institutional Review Board. Research Division. Research involving human subjects. Informed consent for participants of investigative projects. Blacksburg: Ed. Virginia Polytechnic Institute, 1995.
26. Aracil Rodríguez E, Banegas Banegas JR, Bengoechea Bartolomé E, Del Llano Señarís J, De la Mata Barranco I, Gol Greixa J et al. Sistema gráfico de información sanitaria en España (2.<sup>a</sup> ed.). Madrid: Artursa, 1996; 59.
27. Bartek PA. Increasing subject comprehension of the informed consent form. *Drug Inform J* 1995; 29: 91-98.
28. Simón P, Concheiro L. El consentimiento informado: teoría y práctica (I). *Med Clin (Barc)* 1993; 100: 659-663.
29. Meisel A, Kuczewski M. Legal and ethical myths about informed consent. *Arch Intern Med* 1996; 156: 2.521-2.526.
30. Mattson ME, Curb JD, McArdle R, and the AMIS and BHAT Research Groups. Participation in a clinical trial: the patients' point of view. *Controll Clin Trials* 1985; 6: 156-167.
31. Marshall S. Participants should be given feedback about the trial [carta]. *Br Med J* 1996; 312: 186.

Parte 6

---

**COMITÉS DE ÉTICA**

---

---

# Comités éticos de investigación: una complicación necesaria

---

María del Mar López Martín<sup>a</sup> y José Emilio Batista-Miranda<sup>b</sup>

<sup>a</sup>Licenciada en Derecho. Unidad de Investigación. Hospital General Universitario de Elche  
<sup>b</sup>Servicio de Urología. Boston University Medical Center.

Recientemente hemos tenido ocasión de observar de cerca el funcionamiento de un comité ético de investigación (CEI) de un hospital americano desde dos puntos de vista complementarios: como investigador que solicita la aprobación de un proyecto (JEB) y como miembro lego visitante en el mismo comité y en otros similares en la misma ciudad (MML). Ello nos ha permitido conocer un poco su funcionamiento diario y su dimensión en un hospital universitario. Quizás el rasgo más importante es que los CEI no están restringidos a los nuevos protocolos o estudios que se lleven a cabo sobre pacientes, sino que revisan cualquier investigación o trabajo que suponga una modificación sobre las prácticas habituales del hospital, nuevas técnicas a aplicar sobre muestras o tejidos, e incluso cuestionarios a suministrar a los pacientes.

Los avances en la ciencia biomédica, y su aplicación en la clínica, han enfrentado a nuestra sociedad con nuevas y difíciles cuestiones éticas. Existe en nuestro país la nueva normativa sobre ensayos clínicos que establece la regulación de los comités éticos de investigación clínica (Real Decreto 561/93 artículos 39 a 43)<sup>1</sup>. Quizá la trascendencia de esa normativa ha pasado desapercibida y su desarrollo completo y aplicación sistemática podría significar un cambio importante en la forma de realizar la investigación. El análisis detallado de esta normativa escapa al contenido de este artículo, pero las similitudes con el sistema norteamericano son importantes.

Por ello creemos que nuestra pequeña experiencia puede ser útil en donde se pretenda poner en marcha un CEI. Dejamos también de lado los comités de ética asistenciales, cuya finalidad es el análisis de los problemas de bioética en el

funcionamiento diario del hospital<sup>2</sup>. En los hospitales deberían existir ambos (comité de ética asistencial y CEI). De hecho el Real Decreto requiere la presencia y conformidad de un miembro del comité de ética asistencial para que la aprobación que el CEI da a un estudio o protocolo sea válida. Por tanto, los CEI se encargan exclusivamente de aprobar los nuevos estudios y protocolos de investigación.

---

## Un poco de historia

La historia de los CEI va ligada al desarrollo de la bioética después de la Segunda Guerra Mundial<sup>3</sup>. El riesgo de ser repetitivos nos parece menor que el riesgo de olvidar episodios poco edificantes en la investigación biomédica. Los CEI, conocidos en EE.UU. bajo la denominación genérica de Institutional Review Board (IRB, textualmente, Comités Institucionales de Revisión) entraron en vigor en el año 1974 cuando el Congreso Norteamericano estableció que estos comités debían evaluar la investigación sobre sujetos humanos financiada con fondos federales. Esta regulación fue consecuencia de la difusión en la prensa de que se habían realizado investigaciones en fetos humanos con el soporte financiero del National Institute of Health. En años anteriores se habían producido las primeras denuncias sobre investigación éticamente cuestionable<sup>4</sup>; el experimento llevado en el Jewish Chronic Diseases' Hospital, donde se inyectaron por vía subcutánea células de cáncer a pacientes, o los estudios del Wilbrowbrook State Hospital, en el que niños con deficiencias fueron deliberadamente infectados con el virus de la hepatitis B<sup>5</sup>. En ambos casos, la falta de consentimiento de los pacientes fue flagrante. El Congreso Norteamericano nombró por Ley 93-348, una Comisión para la protección de los seres humanos (Commission for the Protection of Human Subjects of Biomedical and Behavioral Research), para establecer los

---

Publicado en *Med Clin (Barc)* 1995; 105 (8): 298-301.  
J.E. Batista estuvo becado por la Fundación "La Caixa" (Convocatoria 1994).

principios éticos básicos sobre los que debía conducirse la investigación biomédica. La Comisión desarrolló sus trabajos entre 1974 y 1978, estableciendo unas recomendaciones que identificaban los principios éticos básicos: *respeto por las personas, beneficencia y justicia*. Estos principios deberían conducir la investigación en seres humanos y se agruparon en “The Belmont Report: Ethical Principles and Guidelines for the protection of Human Subjects of Research”, documento que sigue siendo un punto de referencia para la investigación en seres humanos<sup>6,7</sup>. Ello se debe a que el sistema americano siempre ha considerado los códigos internacionales (p. ej., la Declaración de Helsinki) como principios generales, aunque no describan los procedimientos específicos. El Consejo Internacional del Organismo de Ciencias Médicas (CIMOS) ha propuesto en 1993 un acercamiento a este modelo en la investigación en seres humanos aprobando unas directrices que concretan estos principios<sup>8</sup>.

### Aspectos generales

Según la nueva normativa española, los comités deberán evaluar los aspectos metodológicos, éticos y legales de los ensayos clínicos y una de sus principales funciones es asegurar la protección de los derechos de las personas que intervienen en los estudios, a través del llamado “consentimiento informado en investigación” (artículo 12)<sup>1</sup>. El consentimiento informado es la formalización escrita de los *riesgos y beneficios* a que pueden ser expuestos los individuos que participan como objetos de la investigación e implica su comprensión completa y la aceptación expresa por parte de los sujetos de dichos riesgos.

La finalidad del CEI en EE.UU. es la de velar por los derechos y el bienestar de los seres humanos que participan en el proyecto de investigación clínica, de acuerdo con la legislación vigente. También tienen la responsabilidad y competencia de revisar, aprobar, desaprobar, requerir cambios, suspender o dar por finalizada cualquiera de estas actividades investigadoras. La presentación de un protocolo para su aprobación por el CEI es una garantía de que los derechos federales e institucionales serán respetados. Cada CEI establece anualmente un calendario de trabajo, donde se recogen las fechas de sus reuniones y los plazos de presentación de los protocolos<sup>9</sup>. En la mayoría de instituciones visitadas, la periodicidad de las reuniones suele ser mensual, a lo largo de todo el año. Como norma general, el investigador no asiste a la reunión, si bien se requiere su disponibilidad física o telefónica duran-

te la misma, para que pueda aclarar las dudas que puedan surgir en la discusión de su proyecto. De los comités visitados, únicamente el Dana Farber Cancer Institute requiere que el investigador principal acuda a la reunión para explicar brevemente su proyecto y contestar a las preguntas de sus miembros; concluida esta presentación, el investigador abandona la reunión, y el comité procede a la votación del proyecto.

La composición de los miembros del CEI está recogida en la regulación<sup>10</sup> y detalla algunos de los requisitos que éstos deben cumplir. Se requiere un mínimo de 5 miembros, si bien su composición suele oscilar entre los 20 y los 30. Pertenecen a distintas profesiones, principalmente médicos e investigadores que representen a los diferentes departamentos del hospital, enfermería, abogados, miembros legos (hasta 5 miembros en algunas instituciones), asistentes sociales, etc. Además, el CEI puede invitar a otras personas con especial competencia en un área con el fin de colaborar en la revisión de un determinado estudio, sin derecho a voto. Es también costumbre en los hospitales universitarios la asistencia a las reuniones de estudiantes que estén en relación con esta área. La selección de los miembros es llevada a cabo, bien por el director de la Institución o por el presidente del CEI con la idea de que estén representados todos los departamentos. Normalmente, la participación es voluntaria. La pertenencia a los comités es un hecho aceptado como parte del trabajo en la institución y tiene su reflejo en la vida profesional. El período de pertenencia al CEI es de 2 años y puede ser renovado.

El presidente suele ser una persona de reconocido prestigio en la institución y con capacidad para llevar a cabo sus funciones de manera imparcial. Más que ningún otro miembro, tiene como objetivo orientar el común entendimiento, el sentir general de la normativa institucional y el marco ético que conlleva el proceso de revisión. Es responsable de comunicar las decisiones del comité a los investigadores y a la administración de la Institución, respondiendo en nombre de la misma cuando se soliciten informes por el Gobierno Federal. La regulación de la FDA (Food and Drug Administration) establece, así mismo, que el presidente informe directamente de cualquier incumplimiento de la normativa vigente<sup>11</sup>. Tanto las deliberaciones y decisiones del comité, como la información asociada a los proyectos de investigación son de carácter confidencial. La no observancia de la confidencialidad es una de las causas que puede llevar a la separación de un miembro del CEI. Ningún miembro del CEI podrá

participar en la revisión de un proyecto en el que tenga interés, excepto para proporcionar la información necesaria. La relación con los nombres de los componentes, así como los datos referentes a su situación actual, se emiten a una central federal (Office for Protection from Research Risks) y debe actualizarse de modo permanente.

Es importante que el CEI esté dotado de unos medios personales y materiales mínimos: suele existir un coordinador que sirve de enlace entre los investigadores y el CEI. Dispone de toda la información administrativa (requisitos, fechas de presentación, modelos de consentimiento) y suele ayudar a los investigadores cuando surgen dudas en la redacción tanto de la memoria como del consentimiento informado. También se encarga de archivar, copiar y enviar los proyectos a los miembros del CEI. El coordinador suele tener una dedicación completa y la ayuda de otras personas con funciones administrativas (bien a tiempo parcial o estudiantes de ciencias de la salud o derecho). Su titulación depende del volumen de trabajo y de los centros, pero en muchos casos no ha cursado estudios sanitarios. Suele ser miembro (lego o no según sus estudios) con derecho a voto. Su trabajo es fundamental pues ayuda a los investigadores a redactar de forma comprensible tanto el protocolo como el consentimiento informado. Sólo después de una revisión por el coordinador el proyecto pasa a la reunión mensual del CEI. Las coordinadoras de los hospitales visitados (todas mujeres) clasifican los proyectos en función del riesgo que suponen para el individuo. Tanto la figura del coordinador como los medios materiales que precisa (administrativos, de espacio físico) son esenciales para el funcionamiento de un CEI. Si se pretende que un CEI funcione a base de esfuerzo extraordinario de sus miembros y no existe la figura del coordinador de una dedicación prioritaria, dicho comité estará abocado al fracaso. Sirva como ejemplo que en uno de los centros visitados (Brigham and Women's Hospital) el CEI aprobó en 1993 2.260 protocolos y proyectos.

### **Revisión de un protocolo por el CEI**

Existen varias normas éticas aceptadas en los códigos para la evaluación de un protocolo, entre las que cabría destacar:

1. Diseño del protocolo. Una de las causas que justifican la investigación en seres humanos desde el punto de vista ético es el valor social de un avance científico. Por lo tanto, la legislación federal requiere al menos que la determinación del

riesgo a que los sujetos van a ser sometidos esté en relación con la importancia de los resultados que es previsible obtener. Algunos CEI tienen establecidas unas recomendaciones básicas para elaborar los protocolos, ya que la evaluación de los riesgos/beneficios por el comité depende básicamente del alcance y calidad de la información suministrada por el investigador<sup>12</sup>.

2. Competencia del investigador. Tanto el código de Nüremberg como la Declaración de Helsinki<sup>13</sup> hacían ya referencia directa a la competencia del investigador, tanto desde el punto de vista de una experiencia como desde el de su capacidad para llevar a cabo la propuesta.

3. Análisis de la relación riesgo/beneficio. Este apartado está desarrollado en todos los códigos éticos que se refieren a la investigación humana y es asimismo parte importante de la regulación federal. Según Levine<sup>13</sup>, los riesgos para los sujetos del estudio pueden ser físicos, psicológicos, sociales y económicos, pudiendo producir además una invasión de la intimidad, falta de confidencialidad o atentar contra la dignidad de la persona. La mayoría de las instituciones han establecido unas categorías de riesgos, comprendiendo la I aquel excluido de revisión por el CEI, la II el riesgo considerado "mínimo" y la III todo aquel no comprendido en las categorías anteriores (tabla I). En la regulación del Departamento de Sanidad Estadounidense se define como "riesgo mínimo" aquel cuya probabilidad de concurrencia y magnitud es similar a la que encontramos en la vida diaria, o durante la realización de un examen físico o cuestionario psicológico. El análisis del "riesgo mínimo" se determina a partir de otros elementos, tales como la competencia del investigador, la presencia de personal experimentado que realice una vigilancia frecuente de la investigación y la exclusión de personas pertenecientes a grupos con determinados riesgos (mujeres embarazadas y niños). El beneficio que haya de obtenerse del proyecto científico debe ser real y posible, si bien no existe una fórmula que pueda ser aplicada de modo universal para establecer un balance entre riesgo y beneficio.

4. Selección equitativa de los sujetos. Este apartado trata de evitar que ciertas clases de sujetos sean seleccionados sistemáticamente a causa de su mayor disponibilidad (presos, estudiantes de medicina), aunque un celo excesivo en esta norma puede llegar a la sobreprotección de los grupos de riesgo antes mencionados, o incluso su marginación de los proyectos. La selección de los sujetos será lo más amplia posible, estableciéndose mecanismos de difusión pública. La FDA ha establecido los puntos que debe



TABLA I  
CATEGORÍAS DE LOS PROTOCOLOS  
SEGÚN EL RIESGO PARA LOS SUJETOS

---

Categoría I: estudio de observación de conductas públicas, prácticas educativas habituales, cuestionarios y entrevistas en los que los sujetos no son identificados (sólo requiere la revisión por parte del coordinador, que puede dar su aprobación)

Categoría II: riesgo similar al que encontramos en la vida diaria. Son todas las pruebas como recogida de pelo, uñas, saliva, orina, sangre en pequeñas cantidades o líquido amniótico tras la rotura de membranas inmediatamente antes del parto. También incluye los estudios que requieren revisar las historias clínicas, datos de laboratorio, etc. (es precisa la revisión del presidente y de uno o 2 evaluadores con conocimientos específicos. La aprobación suele realizarse mediante un procedimiento abreviado)

Categoría III: todos los procedimientos que impliquen riesgos no incluidos en las categorías anteriores. Incluye los ensayos clínicos con nuevos fármacos o técnicas quirúrgicas, procedimientos diagnósticos, y la evaluación de nuevos productos sanitarios, prótesis, etc. (precisa la revisión por parte de todo el CEI)

---

recoger un anuncio de este tipo<sup>14</sup> y la legislación requiere que una copia del mismo se adjunte al protocolo, para asegurar que la información no sea engañosa, especialmente cuando en el estudio intervienen personas con una enfermedad física o mental, o en situaciones de desventaja educacional y/o económica. En algunos estudios se procede al pago de una pequeña cantidad para compensar la pérdida de tiempo, el transporte, etc. El CEI debe asegurarse de que no existan problemas de coerción, y de que la cantidad ofrecida no pueda dar lugar a la participación con ánimo de lucro. Existen dos normas éticas cuya importancia, en nuestra opinión, merece un ulterior estudio: el proceso de obtención del consentimiento informado, y el derecho del paciente a ser compensado por el daño que la investigación puede producirle. El consentimiento informado para un protocolo de investigación será más completo y detallado que el documento para procedimientos habituales, que ya ha sido analizado de forma muy completa en estas páginas<sup>15,16</sup>. De los ensayos clínicos publicados en 3 revistas médicas españolas, el consen-

timiento informado se mencionaba en algo menos de un tercio de ellos, con una tendencia al alza en el período más reciente<sup>17</sup>.

### Métodos de trabajo

---

Para facilitar la tarea de revisar un protocolo se establecen tres categorías ya mencionadas en función del riesgo que asumen los sujetos al participar en el estudio (tabla I). Para las categorías I y II no es necesaria la revisión por todo el comité sino que la realiza el presidente y ocasionalmente 2 miembros más del CEI. La categoría III (riesgo superior al mínimo) es revisada por 2 miembros expertos en el estudio propuesto. Ellos son los encargados de describir el estudio a los miembros del comité. Sirva como ejemplo el estudio presentado por uno de nosotros (JEB): la realización de un estudio urodinámico en pacientes antes y después de la colocación de una bomba intratecal de baclofeno. Dicho estudio sólo requiere sondaje vesical temporal, pero no es rutinario en dichos enfermos, por lo que se consideró categoría III.

Después de su discusión el presidente propone la votación para la *aprobación, aprobación provisional, aplazamiento o desaprobación*. De esta votación se decidirá el destino final del protocolo. Cualquiera de estas decisiones será notificada por escrito al investigador al día siguiente de la reunión. Si se trata de una *aproximación provisional*, significa que el CEI está básicamente de acuerdo con el diseño del estudio pero solicita algunos cambios o clarificaciones antes de su aprobación. El *aplazamiento* implica que el proyecto queda bloqueado hasta que se lleven a cabo los cambios que garanticen su viabilidad. Su *desaprobación* por el CEI supone que se han encontrado problemas referentes al documento del consentimiento informado, confidencialidad, diseño del protocolo o datos insuficientes. Tanto en el caso del aplazamiento como en el de la desaprobación del protocolo, el investigador dispone de 3 meses para solucionar los defectos de fondo o forma indicados por el comité a fin de obtener la aprobación del proyecto. Dicha aprobación no es *en ningún caso definitiva*. Normalmente cuenta con una periodicidad de *un año*, pudiendo ser solicitada su revisión por el comité en un plazo inferior, en relación al grado de riesgo inherente al estudio. La *aprobación provisional* suele ser la clasificación más frecuente y los cambios que se requieren suelen referirse al texto de consentimiento informado, para hacerlo más claro a los sujetos participantes. Un mes antes de la fecha prevista para la revisión del proto-

colo, el investigador tendrá que enviar un informe detallando su estado, a partir del cual el CEI decidirá sobre su actividad, inactividad, su finalización o, en su caso, la suspensión del mismo, cuando se hayan producido resultados distintos a los esperados.

### ¿Una complicación más para la práctica clínica?

Los trámites precisos para la actuación del CEI no son complicados, como tampoco lo son las líneas que definen su composición, reunión o sus métodos de trabajo. Gracias a una estructura sencilla como la que hemos observado, existe una cierta garantía para los pacientes, y un autocontrol de los clínicos e investigadores. No se puede plantear ningún nuevo estudio clínico sin antes tener la aprobación del CEI, norma que siguen voluntariamente la mayoría de los miembros de hospital. Ello obliga a un planteamiento teórico antes de empezar cualquier estudio en vez de lanzarse a una revisión indiscriminada de historias clínicas o a añadir exploraciones a los pacientes sin una idea muy clara del objetivo final. Para algunos clínicos supone una complicación preparar una memoria razonada y estructurada entre la idea y el inicio del estudio. Sin embargo, implica que tendrá que revisar y resumir su proyecto y que está recibiendo un cierto apoyo institucional. El CEI funciona de una manera ágil, abierta y consistente: sus normas son públicas y se conoce a principio de año todas las fechas de las reuniones. Ello contribuye a que muchos clínicos lo consideren como algo positivo. Las modificaciones de los miembros del CEI, sin ningún interés personal en el asunto, dan una dimensión más objetiva al proyecto. Esto es especialmente importante a la hora de conseguir un texto comprensible en el consentimiento informado. Claramente supone una complicación para la estructura de algunos centros sanitarios, en los que muchos comités se puede asociar a "reunión inútil" y provocan rechazo inmediato. Uno de los puntos claves del CEI es la ayuda que encuentra el investigador por parte del coordinador a la hora de elaborar el proyecto. Las ventajas para el hospital o centro sanitario son evidentes: si se consigue que todos los proyectos de investigación sean revisados por el CEI, existe una garantía de que van a tener una calidad mínima, tanto ética como científica. También se estimula la colaboración entre distintos departamentos y se consigue un mayor respeto hacia el paciente.

Aunque estos comités requieren que los investigadores justifiquen su actividad en seres huma-

nos, eso no significa que realicen una investigación necesariamente "ética"; pero al menos los CEI proporcionan un soporte legal y ético a experimentos que de otra manera no podrían realizarse debido al impacto que tendrían en los seres humanos (p. ej., la experimentación para la implantación de un corazón artificial). Pueden aprobarse estudios que luego parecerán cuestionables. Estos "fallos" de un CEI suelen deberse a una excesiva identificación con las necesidades e intereses de la institución y a la falta de experiencia en casos muy complejos desde el punto de vista científico<sup>12</sup>. Por otro lado, la aprobación de un protocolo por el CEI no significa una impunidad total para el investigador, aunque sí un dato a favor suyo en caso de conflictos. Recientemente los tribunales fallaron en contra de uno de los hospitales visitados por aplicar un protocolo quirúrgico experimental a pesar de que había sido aprobado por el CEI. El impacto en la opinión pública fue considerable. No tener la aprobación arroja grandes dudas sobre la transparencia de una actividad clínica o de investigación que se sale de las normas comunes. Las agencias federales (NIH y FDA) establecen al menos anualmente unas reuniones de trabajo sobre los CEI que están abiertas a todas las personas interesadas en la investigación en seres humanos. Tanto las personas que trabajan en los CEI como sus miembros se benefician de estas sesiones.

Los CEI suponen una complicación sobre la actividad que se desarrolla en muchos centros, pero su finalidad es proteger a las personas. Y múltiples ejemplos sugieren que esta protección es necesaria mientras se producen avances científicos que mejoran nuestros conocimientos y posibilidades terapéuticas sobre las enfermedades.

### Agradecimiento

Al Dr. Michael Grodin su ayuda para acceder al programa "Law, Medicine and Ethics" en el Departamento de Derecho y Medicina de la Universidad de Boston.

### BIBLIOGRAFÍA

1. Ministerio de Sanidad y Consumo. Real Decreto 561/1993, de 16 de Abril. BOE n.º 114. 13 de mayo de 1993; 14.136-14.364.
2. Broggi-Trias MA. Comités de ética hospitalarios. *Med Clin (Barc)* 1991; 96: 617-718.
3. Annas GJ, Grodin MA. *The Nazi Doctors and the Nuremberg Code. Human Rights in Human Experimentation*. Oxford: Oxford University Press, 1992.

4. Beecher H, Ethics and clinical research. *N Engl J Med* 1966; 274: 1.354-1.360.
5. Katz J. *Experimentation with human beings*. Nueva York: Russell Sage Foundation, 1972.
6. Galende I. Problemas éticos de la utilización de seres humanos en investigación biomédica. Consecuencias sobre la prescripción de medicamentos. *Med Clin (Barc)* 1993; 101: 20-23.
7. The Belmont Report. April 18, 1979. OPRR Protecting Human Research Subjects. Washington: US Government Printing Office, 1993.
8. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva. Council for International Organizations of Medical Sciences (CIOMS). Ginebra: OMS, 1993.
9. Grodin MA, Zaharoff BE, Kaminow PV. A 12-year audit of IRB decisions. *Quality Rev Bull* 1986; 12: 82-86.
10. Ryan MK. General Organization of the IRB. En: Greenwald RA, Ryan MK, Mulvihill JE, editores. *A handbook for IRB*. Nueva York: Plenum Press, 1982; 29-38.
11. Sección 46.116 Code of Federal Regulations, Title 45 Part. 46. Protection of Human Subjects. Effective August 1991; 19.
12. *Guidelines for Investigators*. Trustees of Health and Hospitals of the City of Boston, Inc. Boston; 1986.
13. Levine RJ. *Ethics and Regulations of Clinical Research* (2ª ed.). New Heaven, Londres: Yale University Press, 1988.
14. FADA. IRB Information Sheet. Advertising for study subjects, 1989.
15. Simón-Lorda P, Concheiro-Carro L. El consentimiento informado: teoría y práctica (II). *Med Clin (Barc)* 1993; 100: 659-663.
16. Simón-Lorda P, Concheiro-Carro L. El consentimiento informado: teoría y práctica (I). *Med Clin (Barc)* 1993; 101: 174-182.
17. Soto J, Galeande I, Sacristán JA. Calidad de los ensayos publicados en España: valoración a través del análisis de 3 revistas durante el período 1985-1991. *Med Clin (Barc)* 1994; 102: 241-245.

---

# Comités éticos de investigación clínica: nuevas funciones para un nuevo escenario

---

María Isabel Lucena, Gianni Tognoni<sup>a</sup> y Felipe Sánchez de la Cuesta

Servicio de Farmacología Clínica. Hospital Clínico Universitario. Facultad de Medicina. Universidad de Málaga.

<sup>a</sup>Laboratorio de Farmacología Clínica. Instituto di Ricerche Farmacologiche Mario Negri. Milán. Italia.

Pocos encontrarían injustificado afirmar que la investigación es esencial en cualquier estrategia para mejorar la salud, configurándose además como un elemento clave para el desarrollo de cualquier especialidad médica y como parámetro de la calidad de la asistencia sanitaria que se ofrece. El médico debe sentir la investigación como el fermento necesario para el desarrollo profesional<sup>1</sup>. Parcialmente superada, pues, en los últimos años una fase de debate teórico en nuestro país para situar la investigación clínica en un contexto y una proporción determinados y con una asignación creciente de recursos económicos públicos y privados, el debate se ha trasladado a cuestiones de ordenamiento orientadas a asegurar, de una parte, el escrupuloso respeto a las normas de buena práctica clínica (BPC) y, de otra, a la auténtica promoción de la investigación de la calidad.

La metodología experimental es el estándar obligado para producir conocimiento y dotar a cualquier propuesta de solidez científica. La amplia aceptación de este principio está suponiendo un considerable incremento en la presentación de proyectos de investigación en los hospitales, fundamentalmente de ensayos clínicos (EC)<sup>2</sup>, mayoritariamente promovidos por la industria farmacéutica<sup>3</sup>, con el objetivo habitualmente de confirmar la eficacia de un medicamento o técnica en una indicación concreta. En otras ocasiones, los proyectos de investigación están respaldados por un grupo de trabajo con un determinado nivel de autoridad en la materia o por sociedades científicas.

No existe mejor método para determinar la eficacia y la seguridad de un nuevo fármaco y que proporcione mayor validez interna que el EC controlado y aleatorio, siendo considerado el estándar obligado. No obstante, requiere un minucio-

so diseño y planificación para que sea científicamente válido y clínicamente relevante, y adolece de ciertas limitaciones, fundamentalmente en lo que respecta a la generalización de los resultados que puedan impedir su extrapolación a un paciente concreto<sup>4</sup>. En otras circunstancias no es posible su realización, por lo que tendríamos que plantear otro tipo de estudios observacionales analíticos (cohortes y casos y controles), o bien desarrollar<sup>5</sup> la lógica y la práctica del ensayo  $n = 1$ . Estos diversos enfoques son, en definitiva, un instrumento para el progreso en la investigación y la evaluación de fármacos o acciones terapéuticas.

Con la normativa existente en la actualidad, es francamente difícil que un investigador pueda constituirse como promotor de un EC debido a las exigencias de disponibilidad de tiempo y económicas requeridas para su realización<sup>6</sup>. De este modo, la iniciativa de la investigación en medicina corre un cierto riesgo de transferirse desde los médicos a la industria privada. En este último escenario, no es infrecuente la confusión de los clínicos y la pérdida de perspectiva respecto a la originalidad, oportunidad, ortodoxia metodológica y relevancia científica de estudios en los que finalmente participa.

Por otra parte, el EC ha de seguir las normas establecidas por las autoridades sanitarias y la comunidad científica. El marco legal<sup>7</sup> para la realización de EC lo constituye la Ley 25/1990 del Medicamento, de 20 de diciembre (considerada de tercera generación, ya que tiene como objetivo contribuir a la existencia de medicamentos seguros, eficaces y de calidad), y el Real Decreto 561/1993, de 16 de abril, por el que se establecen los requisitos para la realización de EC con medicamentos y determina las funciones y responsabilidades concretas de los agentes implicados en la realización de los mismos.

En consecuencia, el desarrollo de un fármaco ha adquirido una mayor complejidad legal y ne-

cesita un enfoque multidisciplinario, siendo necesario involucrar a técnicos y científicos de la industria farmacéutica, y haciendo deseable la colaboración entre la investigación académica y la industrial<sup>8,9</sup>, el concurso de los médicos clínicos y los hospitales o de la asistencia primaria<sup>4</sup> donde están los enfermos y, finalmente, de los comités éticos de investigación clínica (CEIC).

### **Papel dominante de la industria farmacéutica**

El aspecto más interesante de la investigación farmacológica es innovar, descubrir y desarrollar fármacos que ofrezcan nuevos mecanismos de acción o beneficios esperados muy significativos, aportando un enfoque original/novedoso en el tratamiento de una enfermedad. No obstante, el interés de la industria farmacéutica es el registro del producto para su utilización en el hombre, y los servicios de salud que son beneficiarios de esta actividad merecen ser también partícipes del proceso del desarrollo clínico de los fármacos.

Por lo general, no se producen nuevos fármacos en los departamentos académicos. Así, puede afirmarse sin complejos que la moderna farmacopea es resultado predominantemente de la actividad de la industria farmacéutica<sup>10</sup>. En este sentido, los datos son abrumadores. El 92 % de las casi 200 nuevas entidades químicas aprobadas por la agencia de regulación de EE.UU. en la década de los ochenta proceden de la industria farmacéutica. Esta circunstancia puede explicar el exiguo número de EC controlados presentados al Fondo de Investigaciones Sanitarias en los últimos años, que constituían sólo un 5 % del total de estudios sometidos a consideración para su aprobación y posterior financiación<sup>11</sup>. De los diseños clínicos revisados en dicho estudio destaca el incremento de los diseños de cohorte (34 %), siendo todavía escaso el número de diseños del tipo casos y controles. No obstante, el escenario ha mejorado, ya que en España durante el período de 1975-1984 en un total de 485 artículos publicados<sup>12</sup> predominaban los trabajos puramente descriptivos (80 %), y no se encontraba ningún caso de diseños del tipo cohortes ni de casos y controles, en tanto que en las revistas extranjeras y durante este mismo período un 7,4 % de los estudios eran del tipo casos y controles, un 9 % de cohortes y un 15 % EC controlados, siendo el resto fundamentalmente descriptivos<sup>11</sup>.

Es tiempo de aceptar la investigación clínica y farmacológica como una parte integral de los servicios de salud (no como una opción extra, sino central) integrada en la naturaleza misma del ser-

vicio<sup>13</sup>, para desarrollar y promocionar estrategias dirigidas a resolver problemas clínicos, asegurar que los medicamentos eficaces sean rápidamente introducidos en el armamentarium terapéutico, sin descuidar cómo se utilizan, y producir un mapa de mejor práctica clínica y efectividad de nuevos tratamientos.

Frente a ello, se constata que mucha de la investigación clínica realizada es de escasa/pobre calidad<sup>14,15</sup>, lo que representa a su vez un problema de enorme dimensión ética por el despilfarro de recursos financieros, humanos cualificados y de innecesaria exposición de los pacientes a la presentación de reacciones adversas, inevitable en este tipo de investigaciones.

El rol predominante de la industria farmacéutica en la investigación farmacológica produce, además, un efecto adicional negativo sobre la investigación independiente al desmotivar su financiación pública. En consecuencia, los servicios asistenciales quedan a merced de la investigación promovida por la industria farmacéutica, que les garantiza una fuente de ingresos y de la que no pueden excluirse. Pero no puede esperarse que ciertos ensayos, como aquellos que comparan terapéuticas alternativas o determinan los fármacos más eficaces y coste-eficientes, sean planteados por la industria farmacéutica. De tal suerte, que necesitamos no perder la perspectiva en una investigación de mayor calidad y enfocada en los aspectos relevantes<sup>16</sup>.

### **Novedad frente a equivalencia**

Un análisis de los 214 nuevos fármacos introducidos en el mercado mundial<sup>17</sup> durante el período 1990-1994 pone de manifiesto una enorme heterogeneidad en el proceso de registro por las diferentes agencias reguladoras de los países más representativos del 60 % de las ventas farmacéuticas globales: EE.UU., Reino Unido, Alemania y Japón, que además disponen de industrias farmacéuticas muy desarrolladas y competitivas.

Esta heterogeneidad es tanto más sorprendente por cuanto se manifiesta en la selección de los productos considerados prioritarios, innovadores (un determinado fármaco, cuyo valor terapéutico se ha demostrado, ofrece al menos una modesta ventaja para el paciente, comparándolo con tratamientos existentes, y/o representa un tratamiento nuevo en un proceso del que no se dispone de alternativa terapéutica) en el momento de lanzamiento en los diferentes países.

La relación entre fármacos innovadores respecto al total registrado es escasa (31 %), pero la calificación de aquello que es innovador es toda-

TABLA I  
CONTRIBUCIÓN DE LOS COMITÉS ÉTICOS A LA INVESTIGACIÓN CLÍNICA

<i>Logros</i>	<i>Limitaciones</i>
Establecer una cultura de revisión externa entre los investigadores	Indolencia y permisividad en el proceso de toma de decisiones
Mejorar la calidad metodológica de los EC presentados	Observatorio "pasivo" de la investigación que se realiza
Esfuerzo de adhesión a las normas de BPC	Relegados al papel de "filtro" que autoriza el EC y después se involucra poco
Mejora de la infraestructura científica	
Apertura de la investigación a la sociedad	
Aflorar la "investigación encubierta"	Riesgo de excesiva comercialización de la propia ciencia
Introducción de criterios de transparencia económica	

BPC: buenas prácticas clínicas; EC: ensayo clínico.

vía motivo de controversia. Ejemplos recientes de discrepancia entre la agencia reguladora y la comunidad científica son fármacos recientemente registrados como el interferón  $\beta$ -1b, la tacrina, el docetaxel y el topotecán<sup>18,19</sup>.

En determinadas categorías terapéuticas es más frecuente la repetición; en un estudio realizado en Inglaterra<sup>20</sup>, donde se analizan las propuestas de autorización de inclusión de nuevos medicamentos en el mercado farmacéutico durante el período 1987-1989, se constata que, de las 118 nuevas propuestas de registro, el 60 % (52 de 86) de los productos semiinnovadores se incluían en una de las siguientes categorías: cardiovascular, sistema nervioso central y antiinfecciosos, mientras que correspondían a estos grupos sólo el 41 % (13 de 32) de los productos realmente innovadores, lo que adicionalmente indica una diversificación de los programas de investigación y desarrollo de la industria farmacéutica.

España no es ajena a esta problemática, como demuestra una revisión reciente de los fármacos autorizados en nuestro territorio<sup>21</sup> durante 1992-1996, donde se constata cómo, en general, los nuevos fármacos introducidos no aportan ventajas terapéuticas importantes (58,4 %), siendo los nuevos fármacos que presentan una contribución excepcional (4 %) o una mejoría terapéutica importante (14 %) destinados para indicaciones poco frecuentes y generalmente de uso hospitalario.

### Comités éticos de investigación clínica

Tal y como se establece en el citado Real Decreto, todos los EC, para poder ser realizados en España, deberán contar con el informe previo del correspondiente CEIC.

La implantación de los CEIC en España, cuyas funciones primordiales son garantizar la validez científica del estudio y el respeto a postulados éticos, salvaguardando los derechos y el bienestar de los sujetos participantes<sup>22</sup>, ha contribuido a (tabla I):

– Establecer una cultura de revisiones externas entre los investigadores. Son comités de revisión que tienen la facultad de aprobar o rechazar un proyecto de investigación. Presupone una toma de decisiones colectiva en el seno de un proceso interdisciplinario abierto, dentro del marco legal y teniendo en cuenta las características locales para llegar a la decisión más acertada y justa. Cabe apuntar, no obstante, una cierta indolencia, permisividad de los CEIC que reflejan tanto factores individuales como elementos locales<sup>23</sup>.

– Mejorar notablemente la calidad metodológica de los protocolos de ensayos clínicos presentados en respuesta a la madurez y experiencia en investigación compartida por administración, comités e investigadores. No obstante, queda aún espacio para el perfeccionamiento<sup>24</sup>. Durante la última década, muchos aspectos éticos de la investigación clínica han mejorado manifiestamente<sup>6,25</sup>, en un esfuerzo de adhesión a las normas de BPC, como son la calidad de la hoja de información al paciente –una forma de abrir la investigación a la sociedad–, el archivo sistemático de la documentación, la información a los pacientes de la existencia de compensación prevista para cubrir los posibles daños y perjuicios que pudieran derivarse de su participación en el ensayo o la comunicación a los CEIC del contrato económico establecido entre investigadores, promotor y centro hospitalario. Además, entre los médicos clínicos ha crecido enormemente el interés por los aspectos metodológicos y técnicos de la in-

vestigación. La creciente participación en ensayos multicéntricos internacionales es reflejo de la mejora de la infraestructura científica de muchos grupos de investigación españoles.

– Un mejor conocimiento de la llamada “investigación encubierta” que se lleva a cabo en los hospitales y la introducción de criterios de transparencia económica. Las relaciones entre investigación científica y sector industrial son estrechas, existiendo una necesaria transferencia de tecnología entre el nivel académico y el productivo que puede conllevar el riesgo de “excesiva comercialización de la propia ciencia”<sup>26</sup>. Los CEIC, al tener la responsabilidad de conocer las relaciones económicas derivadas de la realización de los ensayos clínicos, se han convertido en un primer foro de cultura y sensibilización de los investigadores sobre esta problemática.

El ordenamiento del proceso de evaluación y registro de medicamentos en un único ámbito europeo va a forzar a los CEIC a actuar al tiempo con gran eficacia, pero de forma competitiva<sup>27</sup>. Un ejemplo de los nuevos desafíos que tienen que acometer es la revisión de un protocolo de un estudio multicéntrico por los diferentes CEIC de los hospitales participantes, que lejos de incrementar su calidad supone más bien un entolecimiento en el proceso de aprobación si se realiza alguna modificación, cuando no el propio carácter multicéntrico del estudio “presiona” indirectamente al CEIC obligándole a realizar una valoración rápida del mismo para no “quedarse atrás” respecto a otros centros hospitalarios en los que pudiera estar en curso.

Paradójicamente, el incremento de la responsabilidad y autonomía de los CEIC, que supone aspectos tan importantes como la protección de los derechos, seguridad y bienestar de los sujetos participantes y requiere un seguimiento del EC desde su inicio hasta la recepción del informe final, para asegurar que la investigación se realiza según los protocolos aprobados, no ha tenido reflejo en la práctica, permaneciendo a menudo como observatorios pasivos de la investigación que se realiza en el centro y relegados al papel de “filtro” que autoriza, pero que a partir de este momento se involucra muy poco en la realización del mismo.

### *Distancia entre principios y hechos*

Aunque los objetivos primordiales de la investigación médica han sido intentar comprender los mecanismos de la enfermedad y producir nuevos tratamientos, el análisis riguroso de la efectividad de los nuevos tratamientos ha sido relegado a un

lugar secundario. En consecuencia, muchos tratamientos nuevos han tardado años en alcanzar la práctica clínica, otros ineficaces se han utilizado ampliamente y la medicina se ha basado más en la opinión que en la evidencia científica<sup>13</sup>.

Es una realidad patente que provoca legítima preocupación el elevado número de fármacos comercializados, su excesivo consumo y la gran distancia existente en muchas ocasiones entre los EC controlados y el uso clínico en la práctica diaria<sup>28</sup>. Es, además, fácilmente constatable la tendencia relativamente creciente a colocar productos en el mercado farmacéutico que no han sido suficientemente evaluados, bien porque se han utilizado diseños simples con objetivos discutibles o poco ambiciosos o porque no se han seleccionado correctamente las variables sobre eficacia y, aunque han sido correctamente planificados o ejecutados, no ofrecen las respuestas necesarias<sup>29,30</sup>.

En otras ocasiones, un tamaño muestral inadecuado o una duración limitada del estudio<sup>31</sup> impiden contestar dichas preguntas. Un EC así diseñado carecería, además, de validez ética, ya que situaría a los pacientes participantes en una posición de riesgo injustificado y la extrapolación de sus resultados perjudicaría a poblaciones susceptibles, al prescribírselos tratamientos ineficaces o peligrosos<sup>32</sup>.

Se aprecia, en otras áreas, una discrepancia significativa entre una evidencia experimental altamente prometedora y los hallazgos clínicos contradictorios que son, a pesar de todo, ampliamente aceptados sin haber sido adecuadamente evaluados<sup>33</sup>.

Otro aspecto importante lo constituyen los problemas de extrapolación, de validez externa o capacidad de generalización de los resultados de un EC que se ha llevado a cabo en condiciones de estricto control clínico, sobre poblaciones muy homogéneas<sup>4</sup>. El poder garantizar la representabilidad de los EC es uno de los mayores retos culturales y metodológicos de la farmacología y requiere diseños orientados a examinar no la eficacia, sino la efectividad (rendimiento en condiciones habituales de utilización) de los fármacos. Estos EC ofrecen la ventaja de que escogiendo a la población como objeto directo de la investigación clínica se evalúa el medicamento en las condiciones habituales en las que es prescrito para determinar su impacto epidemiológico en lo que podría ser el “curso natural” de la enfermedad. La idea central es incluir muestras más heterogéneas, pero representativas, de pacientes (lo que exige un número muy elevado) para poder poner de manifiesto efectos benefi-

ciosos, aunque “moderados”, de los tratamientos sobre parámetros de gran relevancia clínica. Al ser dotados los EC de un armazón epidemiológico, tienden a parecerse a una “vigilancia epidemiológica”, a la vez que pueden ser contemplados como un programa educativo permanente. Las fronteras entre investigación clínica y epidemiología se hacen de esta forma menos evidentes<sup>34</sup>.

Otro motivo de preocupación es la constatación de que el flujo de productos “nuevos” de menor calado innovativo<sup>21</sup> no ha disminuido a pesar de las estrategias destinadas a conseguir unos mayores estándares de exigencia en la prescripción médica. Todo ello no hace más que subrayar el hecho desafortunado de que el desarrollo de la investigación clínica está más atento a las exigencias del fármaco y su registro que a la resolución de problemas clinicoepidemiológicos no resueltos.

En todo caso, y en ausencia de evidencia científica, la introducción de una tecnología médica en la práctica clínica es éticamente más aceptable si se realiza en el contexto de un EC controlado y aleatorio que si se prescribe de forma rutinaria a partir de un presumible conocimiento aún no constatado de sus efectos clínicos<sup>28</sup>.

### *Vigilancia de los métodos que producen la evidencia*

El desarrollo de un EC con las exigencias metodológicas actuales requiere la participación de profesionales con vocación investigadora. Su complejidad y carácter interdisciplinario obliga a ensamblar a un grupo heterogéneo y especializado de investigadores, coordinándolos adecuadamente. Sobre todo, exige una minuciosa dedicación y rigurosa metódica en la recogida y análisis de los datos. Monitorizar es un apartado esencial en todo el proceso de producción del conocimiento, pero no debe hacerse con un enfoque reduccionista que el clínico rechazaría –curiosamente el clínico plantea abiertamente un caso difícil en una sesión clínica, pero es habitualmente más reacio a ofrecer detalles sobre la investigación en la que está participando–, sino con la finalidad de sortear las múltiples dificultades que toda investigación genera y hacer germinar al tiempo en el hospital –en demasiadas ocasiones con una visión finalista e inmediata (que aprovecha y aplica los logros de la investigación clínica)– toda una cultura de metodología científica y de responsabilidades compartidas. Esto no puede hacerse si no es desde un cambio en las actitudes y en los comportamientos.

Una implicación práctica inmediata es que los investigadores deben tener una mayor responsabilidad en la planificación y análisis de los protocolos de EC en los que participan.

Tomemos el ejemplo del farmacólogo clínico cuya disciplina combina los conocimientos básicos y la experiencia clínica, ocupándose del estudio científico de los medicamentos en el hombre. Su participación en las tareas del CEIC deriva de las raíces del conocimiento más propio e íntimo de la farmacología clínica. Ya en 1970, la Organización Mundial de la Salud (OMS)<sup>35</sup> establecía como objetivos de formación de farmacólogos clínicos la adquisición de “...los conocimientos metodológicos suficientes para poder planificar, realizar y valorar protocolos de investigación clínica con medicamentos ya comercializados o que se encuentren en fase de desarrollo”, pero que a su vez conduce a enfoques integrados, multidisciplinarios con otras especialidades clínicas para facilitar una investigación de excelencia que haga posible después un uso racional de los fármacos. Los farmacólogos clínicos se han constituido, pues, en una parte integral del proceso de desarrollo de nuevos fármacos<sup>36</sup>, siendo su esencia la valoración/medición de la acción de los fármacos en el hombre y ocupan una posición clave en la promoción en la cultura hospitalaria de todo un proceso de metodología y de investigación, y cuyos resultados van a tener como consecuencia una modificación de la práctica de la medicina en este nivel asistencial. Lo que es aplicable al farmacólogo clínico, puede también serlo al CEIC que se convertiría en observatorio activo/inteligente de la cultura que se desarrolla respecto a los siguientes aspectos (tabla II):

*Evolución/calidad metodológica.* Desde esta perspectiva, el farmacólogo clínico debe asumir el desafío de desarrollar y usar nuevos métodos de evaluación de respuestas de fármacos, traspasando las tradicionales fronteras de los métodos más accesibles en las mediciones de respuesta, y animar la realización de diseños de EC innovadores que aceleren o faciliten (mediante la reducción del tamaño muestral) la obtención de datos sin comprometer la calidad de los mismos<sup>37</sup>. Se han sugerido nuevos diseños en los que “grupos satélites” de pacientes con mayores variaciones pronósticas o “satélites de períodos de tratamiento” con variaciones de regímenes de dosificación mayores sean añadidos al núcleo de un diseño confirmatorio (el enfoque poblacional)<sup>38</sup>.

*Relevancia clínica/epidemiológica.* Favorecer el desarrollo de otras estrategias de intervención



TABLA II  
PROSPECTIVA DE LOS CEIC FARMACÓLOGOS CLÍNICOS COMO "OBSERVATORIOS  
ACTIVOS"

<i>Objetivos</i>	<i>Desafíos</i>
Evolución/calidad metodológica	Nuevos métodos de evaluación Diseños innovadores
Relevancia clínica/epidemiológica	Otras estrategias de intervención: ensayo n = 1, estudios poblacionales, diseños factoriales, análisis caso-control
Exploración de los mecanismos fisiopatológicos de las enfermedades	Trasladar los conocimientos básicos a la práctica clínica
Valorar las propuestas como material didáctico	Debate sobre diferentes tipos de diseño, sesgos de asignación, etc.
Evaluación del impacto del ensayo clínico en las prácticas locales	Valorar la inclusión de fármacos en el vademécum hospitalario
Monitorización del éxito de los protocolos aprobados	El análisis de la investigación realizada, espejo de los límites de la propia investigación
Instrumentos de crecimiento de la investigación	Publicación de la investigación que el CEIC autoriza

CEIC: comité ético de investigación clínica.

farmacológicas o no en aquellas circunstancias donde se estime que la realización de un EC no sería factible. En esta línea, se enmarca el desarrollo de la metodología de los grandes estudios poblacionales, sin interferencia con la asistencia clínica cotidiana, que ha sido uno de los logros fundamentales en las dos últimas décadas<sup>39</sup>. Para resolver problemas complementarios y con características opuestas, se ha desarrollado la lógica del ensayo n = 1. El tercer componente de esta evolución metodológica de la experimentación es la maduración de la estrategia basada en los diseños factoriales y el uso de técnicas de análisis caso-control.

*Intereses de investigación farmacológica, bioquímica, etc.* Aprovechar una investigación farmacológica para descifrar mecanismos causales, usando los fármacos como herramientas para explorar los mecanismos fisiológicos y la fisiopatología de las enfermedades (p. ej., el conocimiento de los mecanismos neurobiológicos que intervienen en el incremento de respuesta de los inhibidores de la recaptación de serotonina al asociar pindolol<sup>40</sup> facilitará el diseño de fármacos antidepresivos más eficaces). En los momentos actuales, existe una necesidad de investigadores bien formados y entrenados que sean capaces de trasladar los conocimientos científicos básicos a la práctica clínica. Esto es particularmente necesario en el área de los productos biotecnológicos.

Los farmacólogos clínicos en el seno del CEIC deben ampliar su enfoque pasando desde los aspectos exclusivamente farmacológicos a aquellos que interesan realmente a los clínicos, entre los que cabe citar a modo de ejemplo el averiguar cómo se modifica la función ventricular, el flujo sanguíneo pulmonar<sup>41</sup> o cuál es la cascada de procesos inflamatorios que suceden durante una sepsis.

*Valoración de las propuestas como material didáctico.* Los EC pueden ser considerados como un valioso material didáctico para fomentar el debate sobre los diferentes tipos de diseño de un estudio (es más difícil explicar un diseño factorial que presentar un diseño de EC con esta aplicación), los posibles sesgos de asignación y los parámetros necesarios para la medida de su calidad, así como el tipo de información que debe facilitarse al paciente.

De otra parte, no puede olvidarse que el objetivo de un EC es comparar la efectividad de un nuevo tratamiento frente a una terapia estándar. No obstante, en un contexto como el actual de progresos médicos en rápida expansión se hace difícil determinar en qué medida las mejoras observadas son la respuesta específica a nuevas terapias y, en consecuencia, las opciones terapéuticas (nuevas y antiguas) no se encuentran suficientemente validadas<sup>42</sup>. El grado de certeza para juzgar una terapéutica como eficaz se sustenta, pues, más sobre los valores de los individuos involucrados que en cálculos científicos.

*Instrumentos de evaluación del impacto del EC en las prácticas locales.* Sopesar los aspectos positivos frente a los negativos que comportan los EC con fármacos semiinnovadores es un delicado desafío futuro de los CEIC. Difícilmente puede objetarse que este tipo de fármacos permite obtener progresos significativos, aunque de escasa magnitud, si bien estas ventajas por lo general se limitan a un número determinado de moléculas de un mismo grupo farmacológico (pero no conseguirán ponerse de manifiesto para la molécula número 15). Como riesgo adicional, la participación en este tipo de EC puede estimular la inclusión en el vademécum personal u hospitalario de un número inaceptable de fármacos semiinnovadores, y se resiente, así, la calidad de la prescripción<sup>45</sup>. En el lado opuesto de la balanza, hay que situar consideraciones como el atractivo financiero que puede suponer para el hospital la contratación de médicos investigadores, adquisición de equipos o material clínico útil, o incluso el suministro gratuito del fármaco durante el período de estudio. Además, este tipo de estudios –por lo general multicéntricos– puede brindar al clínico la oportunidad de revisar críticamente actividades rutinarias (p. ej., plantearse si existen diferencias entre centros, en valores analíticos, criterios endoscópicos o de otra naturaleza) y que de otra forma no se cuestionarían. Se favorece así un clima de diálogo para el desarrollo de una mentalidad experimental, epidemiológica y analítica.

No obstante, la preocupación por el despilfarro de recursos económicos y humanos que comporta este tipo de investigaciones ha llevado a las autoridades sanitarias a hacer un llamamiento a la industria farmacéutica para que oriente sus esfuerzos hacia la obtención de moléculas que representen verdaderos avances terapéuticos, que permitan tratar tanto enfermedades que en la actualidad carecen de respuesta como procesos patológicos poco frecuentes y olvidados por la investigación farmacológica.

*Monitorización del éxito de los protocolos aprobados.* El CEIC comparte, en cierta medida, esta responsabilidad conjuntamente con los coordinadores del EC y con los promotores de la investigación. No es infrecuente que haya dificultades tanto en el reclutamiento de investigadores como de pacientes que, en casos extremos, impiden la realización del estudio, o bien comprometen la validez de los análisis estadísticos efectuados con riesgo de obtener resultados no concluyentes, representando un despilfarro de recursos y un abuso de la buena voluntad de los pacientes<sup>14</sup>. Muchos de estos EC no relevantes son posteriormente

publicados, completando un círculo vicioso del que es difícil imaginar una salida.

Los CEIC deben constituirse en motor de investigación relevante, y eficiente con los recursos económicos y humanos. De tal suerte, que el análisis de la investigación que actualmente se realiza en nuestros hospitales pueda servir de espejo que refleje los límites de la propia investigación.

*Instrumentos de confrontación/crecimiento de la investigación.* Cuando un protocolo de EC es debatido genera diferentes opiniones en el seno de un mismo CEIC y el informe final varía, de hecho, de un comité a otro. Existe, por lo tanto, un componente de variabilidad (que afecta a la calidad) en todo el proceso de toma de decisiones realizado por los CEIC<sup>44</sup>.

Se ha llegado a señalar, en esta línea, que si los CEIC son garantes de los derechos del paciente, ¿quién vigila las observaciones divergentes sobre los aspectos éticos de la investigación médica que se produzcan en su seno? Una forma idónea que tiene el CEIC de someter sus deliberaciones al dictamen de la sociedad (auditoría de lo aprobado) es a través de la publicación de la investigación que, en definitiva, ellos autorizan<sup>45</sup>. Desgraciadamente, esta actitud constituye más la excepción que la norma cuando la insistencia *a priori* por parte del CEIC de publicar los resultados es una responsabilidad esencial que tan sólo puede conducir a mejorar la calidad de la investigación<sup>46</sup>. Actuar de forma contraria es una crítica tanto al tipo de investigación que se realiza como a las motivaciones en las que se sustenta.

### **CEIC: observatorio activo/inteligente de la cultura que se desarrolla**

Dado que los fármacos han dejado de ser las “balas mágicas” del pasado, uno de los principales retos en el desarrollo de medicamentos en el momento presente es la enorme variabilidad en la respuesta motivada por factores cinéticos, dinámicos y genéticos de reacciones adversas o tolerancia. Este hecho motiva la búsqueda de polimorfismos: identificando a los pacientes que no responden a determinados fármacos o desarrollan determinadas enfermedades (p. ej., obesidad), polimorfismos que justifican las variaciones en respuesta farmacológica o requerimientos de dosis<sup>47</sup>.

Otro importante desafío que afronta la farmacología clínica es el desarrollo de productos de biotecnología/genética, cuya especificidad de especie determina que sea menos predictiva de lo habitual, “demandando” más estudios de farma-

cología humana. Estos productos que han supuesto avances significativos para el tratamiento de diversas enfermedades, tales como la vacuna antihepatitis B, la hormona del crecimiento, la eritropoyetina humana, los factores estimulantes, los anticuerpos monoclonales o el interferón alfa, entre otros, requieren, no obstante, de EC con un buen diseño y un particular enfoque en el reconocimiento de reacciones adversas inesperadas.

Si bien es comprensible, en el desarrollo comercial de un fármaco, que se dé una máxima prioridad a la obtención de datos confirmatorios, que justifican su aprobación en el mercado por las agencias reguladoras, la visión del proceso tiene que ampliarse a todo un proceso de aprendizaje de los mecanismos íntimos que pueden justificar en determinadas circunstancias la ausencia de diferencia observada, variaciones en farmacocinética o farmacodinamia en diversos grupos de pacientes o variaciones en regímenes posológicos según factores pronósticos, todo lo que en definitiva constituye una fuerte evidencia de la necesidad de realizar ajustes dosis-respuesta. Este “aprender mientras se confirma” es la estrategia necesaria de desarrollo e investigación de fármacos científicamente orientada<sup>38</sup>.

Así, si tuviéramos que resumir las tres etapas del desarrollo farmacológico, diríamos que la primera se caracterizó por la seguridad, le siguió la era de la efectividad y a ésta una etapa de necesidad de individualización de la terapéutica y dosis-respuesta. En la clínica, sin embargo, y en una dirección totalmente opuesta al intento de individualizar la dosis para contrarrestar la gran variabilidad en la respuesta, se pretende tozadamente unificar al máximo la pauta de dosificación (siendo el paradigma ideal una pauta una vez al día). En este terreno, hay una necesidad evidente de acumular datos que demuestren esta variabilidad y efectuar un cambio en la estrategia terapéutica. Posteriormente, será la propia industria farmacéutica la que se vea forzada a modificar la estrategia. La tradicional falta de atención a estos aspectos ha motivado que ocasionalmente los fármacos se comercialicen a dosis más elevadas (incluso tóxicas) que las que posteriormente se han demostrado eficaces. En sí mismo, los farmacólogos clínicos tienen que ampliar sus intereses y desarrollar/practicar sus conocimientos en una nueva frontera “arena” de su ejercicio dentro de los CEIC.

Los CEIC no pueden permanecer ajenos a este escenario complejo y, en particular, el farmacólogo clínico que ocupa una posición privilegiada para, en íntima interrelación con los clínicos, constituirse en motor de una nueva actitud: “ob-

servatorio activo/inteligente” de la cultura que se desarrolla. Transformando lo que sería una función esencialmente observacional “secretarial” –recogida de reacciones adversas y requisitos legales formales– en una aplicación integral del concepto epidemiológico a la vigilancia sobre culturas y métodos que producen la evidencia.

## BIBLIOGRAFÍA

1. Does research make for better doctors? [editorial]. *Lancet* 1993; 342: 1.063-1.064.
2. Diego Sáiz P, Cuaresma Lasheras E, Tristán Antona C, Chaparro de la Barrera PP. Ensayos clínicos en España (1993-1995). Monografías Técnicas, nº 21. Madrid: Ministerio de Sanidad y Consumo, 1996.
3. Hampton JR, Julián DG. Role of the pharmaceutical industry in major clinical trials. *Lancet* 1987; 2: 1.258-1.259.
4. Lucena González MI. Ensayos clínicos en la asistencia primaria. En: Carné X, Costa I, editores. Problemas y controversias en torno al ensayo clínico. Monografías Dr. Antonio Esteve, nº 22. Barcelona: Fundación Dr. A. Esteve, 1998; 22: 83-93.
5. Cabello López JB, Abraira V, Gómez García J. El ensayo clínico para un solo paciente. Justificación, metodología y aportaciones bioéticas. *Med Clin (Barc)* 1997; 109: 592-598.
6. Espinosa E, Zamora P, Feliu J. Reflexiones sobre la ley de ensayos clínicos. *Med Clin (Barc)* 1996; 106: 24-26.
7. Ley 25/1990 del Medicamento, de 20 de diciembre, Ministerio de Sanidad y Consumo. BOE 1990; 306: 38.228-38.246 (22 de diciembre).
8. Erill S. Investigación académica e investigación industrial en el desarrollo de fármacos. *Med Clin (Barc)* 1987; 89: 247-249.
9. Sykes R. Innovation in the pharmaceutical industry. Partnerships with outside organisations will become imperative. *Br Med J* 1994; 309: 422-423.
10. Dal-Ré R. Coste, eficiencia y próximo futuro del desarrollo de nuevos fármacos. *Med Clin (Barc)* 1997; 108: 621-628.
11. Gómez de la Cámara A. Análisis de la estructura metodológica de las propuestas de investigación presentadas al Fondo de Investigación Sanitaria. *Med Clin (Barc)* 1997; 109: 445-451.
12. Álvarez-Dardet C, Mur P, Gascón E, Nolasco A, Bolumar F. La investigación clínica en España: tipos de diseño utilizados. *Med Clin (Barc)* 1997; 89: 221-223.
13. Smith R. The scientific basis of health services. Needs a better balance between original and evaluative research. *Br Med J* 1995; 331: 961-962.
14. Wise P, Drury M. Pharmaceutical trials in general practice: the first 100 protocols. An audit by the clinical research ethics committee of the Royal College of General Practitioners. *Br Med J* 1996; 313: 1.245-1.248.

15. Winther FO, Hole OP. Scientific quality of clinical research. An analysis of 40 research projects in pharmacology/pharmacotherapy. *Eur J Clin Pharmacol* 1997; 51: 351-354.
16. Altman DG. The scandal of poor medical research. *Br Med J* 1994; 308: 283-284.
17. Kessler DA, Hass AE, Feiden KL, Lumpkin M, Temple R. Approval of new drugs in the United States. Comparison with the United Kingdom, Germany and Japan. *JAMA* 1996; 276: 1.826-1.831.
18. Euromedicines evaluation: the striptease begins [editorial]. *Lancet* 1996; 347: 483.
19. Garattini S. The ethical dilemma of costly new drugs: the case of  $\beta$ -interferon. *Eur J Clin Pharmacol* 1997; 52 (Supl A): 173.
20. Rawlins MD, Jefferys DB. Study of United Kingdom product licence applications containing new active substances, 1987-1989. *Br Med J* 1991; 302: 223-225.
21. García-Vicente JA, Faixedas MI, Condal M. Evaluation of the new drugs that have been authorized in Spain from 1992 to 1996. *Eur J Clin Pharmacol* 1997; 52 (Supl A): 40.
22. Bakke OM, Carné X, García-Alonso F. Ensayos clínicos con medicamentos. Fundamentos básicos, metodología y práctica. Barcelona: Ediciones Doyma, S.A., 1994.
23. De Abajo FJ, Serrano Castro MA. Problemas prácticos de un ensayo clínico (I). En: García Alonso F, Bakke OM, editores. Metodología del ensayo clínico. Monografías Dr. A. Esteve, nº 11. Barcelona: Fundación Dr. Antonio Esteve, 1991; 11-20.
24. Moodie P. The role of local research ethics committee audit: differences between committees. *J Med Ethics* 1996; 22: 78-82.
25. García López JF. Impacto de la normativa legal en la calidad de los ensayos clínicos realizados en España. *Med Clin (Barc)* 1993; 100: 770-777.
26. Cami J. Conflicto de intereses e investigación clínica. *Med Clin (Barc)* 1995; 105: 174-179.
27. Smith T, Moore EJH, Tunstall-Pedoe H. Review by a local medical research ethics committee of the conduct of approved research projects, by examination of patient's case notes, consent forms, and research records and by interview. *Br Med J* 1997; 314: 1.588-1.590.
28. Lau J, Chalmers TC. The rational use of therapeutic drugs in the 21st century. Important lessons from cumulative meta-analyses of randomized control trials. *Int J Technol Assess Health Care* 1995; 11: 509-522.
29. Nicolucci A, Grilli R, Alexanian AA, Apolone G, Torri A, Liberati A. Quality, evolution, and clinical complications of randomized. Controlled trials on the treatment of lung cancer. *JAMA* 1989; 262: 2.101-2.107.
30. Cavaliere D, Scorpiaglione N, Belfiglio M. Quality assessment of randomized clinical trials on medicinal treatment of diabetic neuropathy. *Diab Nutr Metab* 1994; 7: 287-294.
31. Bland JM, Jones DR, Bennett S, Cook DG, Hainess AP, Macfarlane AJ. Is the clinical trial evidence about new drugs statistically adequate? *Br J Clin Pharmacol* 1985; 19: 155-160.
32. Griffiths M. "Underpowered" trials. *Lancet* 1997; 350: 1.406.
33. Di Mascio R, Marchioli R, Tognoni G. From pharmacological promises to controlled clinical trials to meta-analysis and back: the case of nimodipine in cerebrovascular disorders. *Clinical Trials and meta-analysis* 1994; 29: 57-59.
34. Tognoni G, Laporte JR. From clinical trials to drug utilization studies. En: Dukes MNG, editor. Drug utilization studies. Methods and uses. Who Regional Publications European Series 1993; 45: 213-241.
35. OMS. Farmacología clínica: actividades, servicios y enseñanza. Serie de informes técnicos, 446. Ginebra: Organización Mundial de la Salud, 1970.
36. Temple RJ. The clinical pharmacologist in drug regulation: the US perspective. *Br J Clin Pharmacol* 1996; 42: 73-79.
37. Gram L, Smidt D. Innovative designs of controlled clinical trials in epilepsy. *Epilepsia* 1993; 34 (Supl 7): 1-6.
38. Sheiner LB. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* 1997; 61: 275-291.
39. Tognoni G, Fresco C, Maggioni AP, Turazza FM, on behalf of the GISSI investigators. The GISSI Story (1983-1996): a comprehensive review. *J Intervent Cardiol* 1997; 10: 3-28.
40. Perez V, Gilaberte I, Faries D, Alvarez E, Artigas F. Randomized, double-blind, placebo-controlled trial of pindolol in combination with fluoxetine antidepressant treatment. *Lancet* 1997; 349: 1.594-1.597.
41. Gattinoni L, Tognoni G, Brazzi L, Latini R. Ventilation in the prone position. *Lancet* 1997; 350: 815.
42. Lantos J. Ethical issues. How can we distinguish clinical research from innovative therapy? *Am J Pediatr Hematol Oncol* 1994; 16: 72-75.
43. Denig P, Haaijer-Ruskamp FM, Wesseling H, Versluis A. Impact of clinical trials on the adoption of new drugs within a university hospital. *Eur J Clin Pharmacol* 1991; 41: 325-328.
44. Redshaw ME, Harris A, Baum JD. Research ethics committee audit: differences between committees. *J Med Ethics* 1996; 22: 78-82.
45. Pearn J. Publication: an ethical imperative. *Br Med J* 1995; 310: 1.313-1.315.
46. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *Br Med J* 1996; 313: 1.390-1.393.
47. Dollery C. Clinical pharmacology: future prospects for the discipline. *Br J Clin Pharmacol* 1996; 42: 137-141.

Parte 7

---

**PUBLICACIONES BIOMÉDICAS**

---

---

# Cómo elevar la calidad de los ensayos clínicos y de su publicación

---

Bert Spilker

Director of Project Coordination Burroughs Wellcome Co. Research Triangle Park, North Carolina 27709, EE.UU.

La conclusión que se obtiene de las recomendaciones de este artículo es que realizar ensayos clínicos con un diseño deficiente es una falta de ética. La base para esta afirmación radica en que los pacientes se someten a un riesgo si participan en un ensayo del cual no pueden obtenerse datos convincentes. Esta premisa goza de una amplia aceptación y así lo han hecho constar en muchas ocasiones expertos en ética (Levine, 1986)<sup>1</sup>, expertos en metodología en EE.UU. (Thomas Chalmers, comunicación personal) y en Europa (Turner, 1989)<sup>2</sup> y presidentes de comités de ética (Wim Rosinga, European Ethical Review Committee, comunicación personal).

Si se acepta la premisa de que es una falta de ética realizar ensayos clínicos mal diseñados, la primera cuestión a considerar es si alguno, varios, muchos o la mayoría de ensayos clínicos que se realizan actualmente están mal diseñados o no. Aunque es difícil generalizar sobre la calidad de los ensayos clínicos no publicados, está claro que muchos de los publicados están deficitariamente diseñados. En opinión del autor, es responsabilidad de los comités de ética/consejos institucionales de revisión (CIR) el elevar la calidad de los ensayos clínicos mediante el rechazo a tolerar y aprobar protocolos de ensayos clínicos que no estén bien diseñados. Muchos de ellos se aprueban por los comités de ética/CIR que consideran que el ensayo no compromete éticamente el cuidado y tratamiento de los pacientes, es decir, que los pacientes no son perjudicados directamente. Esta actitud es similar a la de revisar los delitos por comisión e ignorar los delitos por omisión. Los pacientes que participan en un ensayo mal diseñado están expuestos a riesgos innecesariamente y

de modo no ético porque los datos obtenidos no convencerán a la mayoría de lectores o pueden llevarlos a conclusiones erróneas.

*Mejorar la calidad de los ensayos clínicos a nivel macro.* Los directores de las revistas constituyen otro grupo importante dotado de poder para influir sobre la calidad del buen diseño del ensayo clínico. Este grupo de personas puede hacer muchas cosas para perfeccionar la calidad de los ensayos clínicos, tanto a nivel macro como micro. Éstas son unas cuantas sugerencias que pueden llevarse a la práctica con un esfuerzo relativamente pequeño:

1. Apoyar el concepto de desarrollo de patrones universales para el diseño, realización, interpretación, extrapolación y publicación de ensayos clínicos.
2. Publicar una lista de comprobación de toda la información que los autores deben incluir en cualquier manuscrito presentado para su publicación. Se creará una lista de comprobación aparte, centrada en el análisis estadístico. Algunas revistas siguen esta práctica.
3. Suministrar a los revisores unas listas de comprobación de información para evaluar los manuscritos y así colaborar en la evaluación de su calidad.
4. Proporcionar una o más listas de comprobación a los autores que han presentado manuscritos, además de la revisión y decisiones sobre la publicación, realizadas por los revisores. Esta lista de comprobación podría ser una copia de las completadas por los revisores.
5. Instituir una columna periódica en la que se comenten los recientes avances, así como los patrones, en el campo de la metodología de los ensayos clínicos. Incluir debates sobre temas y tendencias metodológicas seleccionadas. Muy pocas revistas de amplia difusión ofrecen actualmente a sus lectores información de esta índole.

---

Publicado en *Med Clin (Barc)* 1992; 98 (8): 303-304. Resumen de la ponencia presentada en el simposio sobre "Publicación de ensayos clínicos en revistas médicas", realizado en Madrid (1990) en el seno de la 7ª Conferencia Internacional de Medicina de la Industria Farmacéutica.

El *British Medical Journal* (*Br Med J*) es una notable excepción. Los artículos metodológicos del *Br Med J* son ampliamente citados, lo cual pone de manifiesto lo valiosos que son para las personas interesadas en este campo. Aunque este tipo de información se presenta en la revista *Controlled Clinical Trials*, ésta no tiene el amplio público lector que se beneficiaría de la misma y que está interesado en el área.

6. Apoyar el concepto de ensayos clínicos bien diseñados en editoriales y en otros medios adecuados (p. ej., en simposios).

*Mejorar la calidad de los ensayos clínicos a nivel micro.* Muchas sugerencias específicas con respecto al tipo de información y datos que deberían publicarse en los ensayos clínicos se han presentado con anterioridad (Spilker, 1986)<sup>3</sup>.

Hay dos problemas importantes acerca de la documentación de los ensayos clínicos en la mayoría de la literatura publicada. En primer lugar, en la sección de métodos, se incluyen pocos detalles para que otro investigador pueda repetir el ensayo. La literatura preclínica, en cambio, ofrece por lo general unas descripciones más completas de los métodos utilizados. La mayoría de estudios de bioquímica o farmacología, por ejemplo, se pueden repetir de manera exacta o muy aproximada utilizando únicamente la información publicada en la literatura científica.

En segundo lugar, los relativamente escasos resultados y métodos presentados en la mayoría de publicaciones clínicas no permiten hacer una interpretación adecuada de los datos sobre el ensayo. Una documentación más detallada y completa, particularmente en la sección de resultados, permitiría que se evaluara mucho mejor la calidad y el valor de cada ensayo clínico. Además, los investigadores que quisieran confirmar o ampliar los resultados de los ensayos clínicos publicados verían mejorada la capacidad de diseñar posteriores ensayos. Las comparaciones de los resultados entre ensayos publicados y metaanálisis existentes también se facilitaría si se publicara un mayor número de detalles metodológicos, así como también se facilitaría la preparación de nuevos metaanálisis.

Con una calidad superior de las publicaciones clínicas y el apoyo de los directores de las revistas, se lograría cumplir prontamente estos dos objetivos encaminados a ofrecer datos más completos que permitirían mejores interpretaciones. Los directores de las revistas deben decidir si estos objetivos son importantes y también cómo se debería mejorar la calidad. Los directores de las revistas también deberían presionar a los autores

para elevar la calidad en todos los aspectos del proceso de elaboración del ensayo clínico (es decir, diseño, realización, análisis e interpretación del ensayo). En la publicación de ensayos clínicos, se deberían incluir numerosos temas específicos (aunque, con frecuencia, no lo son). La mayoría de ellos atañen a la información que no consta en la sección de métodos y son los siguientes:

1. Intereses económicos (si existen) de los autores en el resultado del ensayo. Por ejemplo, debería indicarse si el ensayo está realizado por académicos con una poderosa relación económica, con una pequeña compañía de biotecnología que se beneficiará de los resultados del ensayo. Algunas revistas exigen ya esta información.

2. Detalles de cualquier modificación del protocolo en el curso del ensayo clínico.

3. Definición de lo que constituye una respuesta clínica positiva o un paciente que responde. Esto permite que el lector considere y evalúe el criterio de lo que constituye una respuesta clínicamente significativa y qué tipos de respuesta pueden ser solamente estadísticamente significativos.

4. Métodos utilizados para la aleatorización (p. ej., variables de estratificación, diseño de bloques).

5. Métodos de selección de pacientes y si éstos se modificaron durante el ensayo. Si así sucediera, manifestar que se compararon los resultados de cada grupo para garantizar la homogeneidad. En muchos ensayos clínicos, se deberá indicar el número de pacientes con los que se estableció contacto, los que se seleccionaron, los que firmaron consentimientos informados y los que completaron cada parte del ensayo.

6. Los años y meses en que se realizó el ensayo.

7. Una descripción relativamente completa de todos los hechos adversos observados. Raramente se mencionan los descriptores específicos sobre hechos adversos.

8. Grado de cumplimiento alcanzado por los pacientes, más los métodos utilizados para determinar el cumplimiento.

9. Grado alcanzado de "ceguera", dispositivos de salvaguardia para garantizar la integridad del ciego y métodos utilizados para evaluar y validar la calidad de ciego del ensayo.

10. Métodos utilizados en ensayos multicéntricos para garantizar una realización similar del ensayo en lugares distintos (p. ej., mesas redondas, intercambio de investigadores, vigilancia rigurosa,

edición de un manual del investigador, adiestramiento de los coordinadores del ensayo).

11. Métodos utilizados para el procesamiento de los datos y procedimientos utilizados de control de calidad (p. ej., doble entrada más evaluación informatizada de las diferencias, comparación del 10% de los datos brutos totales sobre las formas de recogida de datos con registro informática, visitas locales de auditores autorizados).

12. Otros detalles relevantes del protocolo (p. ej., comité de vigilancia de datos, procedimientos de elaboración y dispensación de medicamentos, criterios de ingreso, gravedad de la enfermedad).

13. Cuando se determinan diferentes objetivos finales en un ensayo, se debería garantizar que en el protocolo se indican los principales que se han utilizado para decidir si el resultado debe considerarse positivo.

14. Muchos de los detalles relacionados con el análisis estadístico que se deberían presentar en una publicación han sido descritos por otros autores (consultar Spilker, 1986, para referencias)<sup>3</sup>.

Se comprende que haya motivos tanto prácticos como económicos que dificulten la puesta en

práctica de estas sugerencias por parte de los directores de revistas. Sin embargo, la mayoría de estas puntualizaciones se puede formular en pocas frases y no son necesarias largas descripciones. Existe una gran tarea que casi todos los directores pueden realizar para elevar la calidad de los artículos publicados. No es que esté a favor de incrementar la extensión de los ensayos clínicos publicados más de un 5 o un 10%. Si se llevaran a la práctica, estas modificaciones incrementarían la atención que los autores e investigadores prestan a estas puntualizaciones. El resultado sería una mejoría no solamente en la calidad de la publicación sino también en el diseño de los ensayos clínicos. Por lo tanto, se diseñaría y realizaría un mayor número de ensayos clínicos sumamente éticos.

## BIBLIOGRAFÍA

1. Levine RI. Ethics and regulation of clinical research. Baltimore: Urban and Schwarzenberg, 1986.
2. Turner P. Clinical trials in elderly subjects. *Postgrad Med J* 1989; 65: 218-220.
3. Spilker B. Guide to clinical interpretation of data. Nueva York: Raven Press, 1986.



---

# La revisión de manuscritos

---

Stephen Lock

Director del British Medical Journal.

La revisión es la evaluación por expertos de un trabajo científico terminado. En el caso de los artículos originales remitidos para su publicación, la revisión se remonta a 1665, fecha de nacimiento de las revistas científicas, cuando los *Philosophical Transactions of the Royal Society* y *Journal des Savans* comenzaron a enviar artículos para su evaluación, inicialmente, a los miembros de las sociedades y, más tarde, a expertos no pertenecientes a dichas sociedades<sup>1</sup>. Esta práctica se formalizó en 1752, cuando la *Royal Society* estableció su *Committee on Papers* (Comité de Artículos). Se sabe que algunas revistas utilizaban dicho comité con regularidad; por ejemplo, en 1893, Ernest Hart, el gran director del *British Medical Journal*, declaró que “cada carta recibida, cada párrafo, cada recorte tratado editorialmente, es enviado a un experto que posee un conocimiento especial y que es una autoridad reconocida en la materia”<sup>2</sup>.

A pesar de estas iniciativas y del hecho que la revisión se convirtió en un requisito legal para obtener una beca de investigación para el *US National Cancer Institute* en 1937<sup>2</sup>, la revisión es principalmente un fenómeno posterior a la Segunda Guerra Mundial; incluso el *Journal of Clinical Investigation*, por ejemplo, no la introdujo hasta 1944<sup>1</sup>. Así pues, aunque probablemente todas las revistas científicas internacionales rigurosas practican la revisión en una forma u otra, su tardía introducción quizá sea una explicación de por qué ha sido tan poco investigada. Han aparecido descripciones aisladas sobre la historia y estudios del proceso de revisión y su resultado, pero para ser un hecho tan crucial para el progreso de la investigación científica, ha sido

asombrosamente descuidado, lo que ha inducido a un orador a manifestar: “lo más pasmoso es la falta de datos sobre el rendimiento del sistema de revisión en vista de la enorme influencia que el sistema ejerce sobre las vidas de las personas que escriben en revistas biomédicas”<sup>3</sup>. En cierto modo, en 1989 se intentó remediar esta negligencia, cuando se celebró la *First World Conference on Peer Review* (que consistió principalmente en presentaciones originales) en Chicago y se prepara otra reunión para 1992.

Los objetivos principales de la revisión son tres: en primer lugar, evitar la publicación de un trabajo de mala calidad (es decir, que el material no sea original o no sea importante o sea de mala calidad científica), en segundo lugar, mejorar el conocimiento y por último mejorar la redacción y la presentación de los datos. Así pues, el artículo, después de una breve revisión inicial en la editorial, es remitido habitualmente a uno o dos revisores a quienes se les encomienda un juicio sobre la originalidad, fiabilidad científica, importancia para la disciplina tratada y adecuación para la revista. Al revisor se le puede solicitar que comente el estilo de redacción y presentación, así como los aspectos estadísticos (o indique si se precisa un examen por un experto en estadística) y que recomiende si el artículo se ha de publicar con modificaciones (especificadas) o sin ellas, o si debe ser rechazado con el consejo de publicarlo en otra revista o rechazado definitivamente. Además, algunos directores solicitan que los revisores asignen a los artículos una puntuación de prioridad para la publicación. Sin embargo, la última decisión la toma el director, aunque, algunas veces, tiene la ayuda adicional de un comité de redacción. La escasa investigación sobre el proceso de revisión ha ofrecido, de un modo sorprendente, resultados decepcionantes acerca de su eficacia. En primer lugar, es costoso; posiblemente cueste, actualmente, unos 200 dólares por artículo a las revistas clínicas importantes ta-

---

Publicado en *Med Clin (Barc)* 1992; 98 (8): 304-305.  
Resumen de la ponencia presentada en el simposio sobre “Publicación de ensayos clínicos en revistas médicas”, realizado en Madrid (1990) en el seno de la 7ª Conferencia Internacional de Medicina de la Industria Farmacéutica.

les como el *New England Journal of Medicine* o el *British Medical Journal*<sup>1</sup>. En segundo lugar, consume mucho tiempo a los expertos; en 1979, éste se estimó en 9 años de trabajo/persona solamente para la primera de las revistas citadas, aun en el caso de que, como se indica en dos trabajos, los revisores tardaran únicamente entre 2 y 2,8 h para evaluar un artículo<sup>4,5</sup>, lo cual tiene como resultado una sorprendente falta de acuerdo, sólo el 41-57 % de una pareja de revisores concordó sobre la decisión de aceptación/rechazo (y una gran parte del consenso atañe a un 25 % de artículos clasificados como malos/impugnables). Y, finalmente, hay tres características deprimentes. En primer lugar, con la insistencia de los autores, el 80-90 % de todos los artículos se publican en alguna revista, con frecuencia sin modificaciones como resultado de los informes previos de los revisores. Además, en segundo lugar, muchos artículos atraen poca o ninguna atención; casi una cuarta parte de ellos nunca se cita como referencia formal. En tercer lugar, en la reciente avalancha de incidentes de inmoralidad o mala conducta científica (piratería, plagio y fraude), la revisión ha demostrado ser sumamente ineficaz para detectarlos.

Además de lo anterior, hay muchos peligros. Uno de ellos es que el revisor puede sufrir un sesgo contra el mensaje del artículo sobre bases personales o filosóficas. Nuevamente, el coste y las demoras en la publicación que supone la revisión de un artículo ¿pueden retrasar el progreso de la ciencia? Y, por último, la timidez del revisor y el director ¿debilitarán o distorsionarán una conclusión importante?

A pesar de todo, un estudio ofrece pruebas de que la revisión es eficaz y que debería continuar. En 1979, tomé 1.588 artículos presentados para su publicación en el *British Medical Journal* y obtuve puntuaciones individuales (en una escala de 1-4) de dos miembros del comité de redacción, revisores y comité de redacción en cuanto a originalidad, fiabilidad científica, importancia clínica, adecuación para publicarse en el *British Medical Journal* y recomendaciones finales sobre su disposición definitiva<sup>1</sup>.

Posteriormente, seguí el destino de los artículos rechazados a través de un examen mediante cuestionario.

Los resultados fueron más satisfactorios de lo que había esperado. Aunque el 68 % de los artículos rechazados por el *British Medical Journal* fue publicado, en un intervalo de 2 años y, con frecuencia, sin modificar, el acuerdo entre los lectores fue superior a la casualidad. A juzgar por el factor impacto de la revista que publicó los ar-

tículos rechazados, el proceso editorial filtró los mejores artículos. Así pues, los artículos rechazados ya por los lectores del comité de redacción se publicaron en revistas de "menor calibre" en comparación con los artículos rechazados en un estadio posterior del proceso (después de la revisión solamente o después de la consideración del comité de dirección). Además, más de la mitad de artículos publicados en el *British Medical Journal* se modificaron como resultado de las sugerencias del revisor o del comité de redacción.

Por tanto, llegué a la conclusión que la revisión era valiosa y que debería seguirse practicando, aunque se ha de mejorar. En primer lugar, cabe considerar lo que necesita el director de la opinión de un revisor. Debe considerar los siguientes puntos: importancia del tema, modo en que se tratan los temas clave, solidez de los métodos y presentación (incluida la estadística). Los autores necesitan una opinión detallada, constructiva, justa, educada y basada en unos conocimientos fundamentados.

También llegué a la conclusión que se podrían introducir algunas mejoras. Si los directores no pueden basarse en un comité de selección para comentar los artículos a intervalos semanales, la mayoría de directores precisa consultar a dos asesores y disminuir el efecto Matthew o efecto halo (las personas eminentes resultan privilegiadas). Estoy a favor de la revisión a ciegas hasta que esté convencido de que es inútil. Los revisores deberán ser rápidos, competentes y justos y deberán ser supervisados frecuentemente por el director y un pequeño comité de redacción. Este último también deberá garantizar que los revisores no reciban una sobrecarga de trabajo. A menos que un revisor lo quiera, no deberá esperarse que revise más de seis artículos por año, dado que la revisión de cada artículo puede llevarle más de medio día de trabajo y, probablemente, recibirá peticiones similares por parte de otras revistas. La elección informatizada de los revisores puede evitar algunos de los sesgos de los directores, lo que les permitirá seleccionar a los menos conocidos así como supervisar su rendimiento.

Las listas de comprobación son útiles para muchos revisores, aunque no para todos, y los directores deberían tratar de mejorar dichas listas de comprobación así como desarrollar otras nuevas. Es necesario comunicar la decisión final a los revisores y enviar una carta explicativa sobre si la decisión final no coincide con su consejo. Se podría dar opción a que los revisores firmaran sus informes si así lo desearan. Se buscará asesoramiento estadístico aparte (informando al estadístico de todos los requerimientos editoriales espe-

cíficos, así como de los informes científicos de los revisores) de todo artículo adecuado que sea aceptado.

Posteriormente, el director leerá nuevamente el artículo a la luz del informe del revisor y decidirá lo que es justo y lo que no lo es, qué modificaciones son practicables y cuáles no y eliminará cualquier frase hiriente antes de enviar el artículo al autor. Aun cuando el artículo sea rechazado, el autor tiene derecho a una explicación: los procesos editoriales y de revisión consumen una gran cantidad de tiempo, gran parte del cual se perderá si se consideran únicamente como unos procesos de clasificación y no, al mismo tiempo, como un ejercicio educativo. Por el contrario, los autores deben aceptar que no todos los rechazos están basados en dudas científicas sobre el artículo sino que pueden reflejar el gusto y el juicio del director (los cuales, si se trata de un buen director, serán tan amplios y generosos como sea posible). Existe también una buena ocasión para introducir algunos mecanismos de llamada de atención cuando los autores de los artículos rechazados puedan ofrecer una oposición fundamentada.

Finalmente, todos los directores deberían proponer o aspirar a algún tipo de auditoría continuada de sus prácticas; cuántas veces recurren a un revisor por año, cuál es la calidad de sus informes y si en un revisor se ha apreciado sesgo, negligencia o errores. El director también debería tratar de seguir los posteriores juicios sobre el artículo que pasa por sus manos, tanto los publicados como los rechazados y publicados en otras revistas.

Esta investigación puede parecer una difícil tarea, particularmente para un director a dedicación parcial con poca ayuda, aunque se facilitaría con un ordenador y la utilización del análisis de citas (particularmente, los nuevos y refinados métodos a medida, tales como los factores de peso); opino que, a pesar de ello, es una tarea esencial. La publicación ocupa una posición tan primordial en la ciencia que es extraordinario que nadie, fuera o dentro de las editoriales, la haya estudiado en la profundidad que su importancia demanda.

Lo más importante, en mi opinión, es la necesidad de aceptar la idea de la revisión a ciegas o, cuando menos, que los directores estudien esta idea en sus propias circunstancias. A las sugerencias previas sobre la fuerza de este argumento, apuntadas por Mahoney<sup>6</sup> y Peters y Ceci<sup>7</sup>, hay que añadir el reciente y riguroso estudio de McNutt et al<sup>8</sup>. En este último, no sólo se demostró que la revisión a ciegas era factible para la edito-

rial y que era un éxito para el 76 % de revisores, sino que, en una escala de 3 puntos, había un 21 % de mejora en la calidad de la revisión, así como un incremento llamativo de la proporción de revisiones excelentes entre los revisores que siguieron este método. De modo que, al igual que urge la repetición de estas observaciones para otras revistas, también urge un estudio del efecto de la ocultación de la identidad de los autores sobre los propios revisores, particularmente, a la vista de otras observaciones recientes<sup>9</sup>. Unos 25 manuscritos que habían sido claramente revisados y aceptados para su publicación en el *Journal of Pediatrics* fueron remitidos para una nueva revisión a otros dos revisores y, posteriormente, a tres directores adjuntos expertos e independientes. Los nuevos revisores opinaron que la mayoría de manuscritos presentaban defectos que justificaban una revisión adicional pero, aunque uno de los directores adjuntos participantes había requerido la revisión con mayor frecuencia que el otro, raramente hubo desacuerdo entre ambos en cuanto a los conceptos básicos de aceptación/rechazo.

Por último, en épocas pasadas, los directores de muchas disciplinas han dormido con la conciencia tranquila independientemente de sus decisiones erróneas a causa del concepto occidental de pluralidad: lo que rechaza *The Lancet* será publicado por *British Medical Journal* o *Gut*. Cicchetti et al<sup>10</sup> han demostrado de modo alarmante que esto no es así para algunas revistas de ciertas disciplinas y para las peticiones de becas de investigación en todas ellas. Se necesitan urgentemente estudios prospectivos y retrospectivos sobre estas observaciones (¿se mantuvieron a través de la historia las ideas rechazadas previamente?, ¿fue el rechazo debido a la falta de rigor de la investigación aunque las ideas originales estuvieran fundamentadas?, ¿fueron estudiadas por alguien más?). ¿Debería ver la sociedad que cierta cantidad de espacio/dinero de las revistas estuviera reservada a ideas "extravagantes" en todas las disciplinas del mismo modo que se fundó el *Medical Hypotheses* de David Horrobin para hacer precisamente eso en biomedicina? Algunos de estos estudios no son demasiado laboriosos y potencialmente serían contribuciones adecuadas a la *Second World Conference on Peer Review*, que se celebrará en 1992 (para detalles, Drummond Rennie, *Journal of the American Medical Association*, 535 N Dearborn St., Chicago, IL 60610, USA).

Algunos fragmentos de este artículo han aparecido en *A Difficult Balance*<sup>1</sup> o en *Behavioral and Brain Sciences*<sup>10</sup>.

## BIBLIOGRAFÍA

---

1. Lock S. A difficult balance. Londres: Nuffield Provincial Hospitals Trust, 1985.
2. Burnham JC. The evolution of editorial peer review. *JAMA* 1990; 263: 1.323-1.329.
3. Ingelfinger RJK. Peer review in biomedical publication. *Am J Med* 1974; 56: 686-692.
4. Lock S, Smith J. What do peer reviewers do? *JAMA* 1990; 263: 1.341-1.343.
5. Yankauer A. Who are the peer reviewers and how much do they review? *Jama* 1990; 263: 1.338-1.340.
6. Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Therapy Research* 1977; 1: 161-175.
7. Peters DP, Ceci SJ. A manuscript masquerade. How well does the review process work? *Science* 1980; 20: 16-19.
8. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. *JAMA* 1990; 263: 1.371-1.376.
9. Garfunkel JM, Ulshen MH, Hamrick HJ, Lawson EE. Problems identified by secondary review of accepted manuscripts. *JAMA* 1990; 263: 1.369-1.371.
10. Cicchetti V et al. *Behav Brain Sci* (en prensa).

---

# La disminución del sesgo en la publicación de los ensayos clínicos

---

Thomas C. Chalmers

Boston Veterans Administration Medical Center. The Technology Assessment Group in the Harvard School of Public Health, Boston, MA y Clinical Trials Unit, Mount Sinai School of Medicine, Nueva York, Nueva York.

La disminución del sesgo en los ensayos clínicos empieza cuando los estudios se diseñan por primera vez. Los investigadores clínicos han de considerar que, como mínimo, hay seis razones por las cuales las conclusiones de los ensayos clínicos pueden discordar (tabla I). Ya sea porque los pacientes en estudio sean lo suficientemente diferentes como para presentar desenlaces diversos, los tratamientos sean lo suficientemente diferentes o que se hayan determinado diferentes objetivos finales. Podría ser que los tratamientos auxiliares fueran tan diferentes que el efecto del fármaco experimental ya no fuera mensurable y, obviamente, las diferencias podrían estar motivadas por tamaños inadecuados de las muestras y por el papel que desempeña el azar. Sin embargo, el punto que debería preocuparnos en esta discusión es la posibilidad de que las diferencias obedezcan a la falta de un adecuado control del sesgo. Este proceso empieza con el diseño del estudio (tabla II).

La selección del diseño de un ensayo clínico que no incluya controles aleatorios es una invitación a una presentación sesgada y errónea. Los investigadores tienen una inclinación natural a seleccionar para un tratamiento nuevo e interesante a los pacientes que responderán pero, en el caso de fármacos potencialmente tóxicos, los investigadores pueden inclinarse a escoger pacientes que no tienen nada que perder, como los afectos de un cáncer terminal.

En ambos casos, es imposible decir si la respuesta del paciente era el resultado del estado de la enfermedad o del efecto de la nueva terapia. Con respecto a la utilización de controles históri-

TABLA I  
DIVERSAS CAUSAS DE SESGO  
Y MÉTODOS PARA SU CONTROL

---

Se han incluido pacientes diferentes
Se han evaluado tratamientos diferentes
Se han medido objetivos diferentes
Se han utilizado tratamientos complementarios diferentes
Las diferencias son el resultado del azar
El sesgo ha influido sobre el resultado de uno o más estudios

---

TABLA II  
OCASIONES PARA EL SESGO EN DISEÑOS  
EXPERIMENTALES DISTINTOS A LOS  
ENSAYOS CONTROLADOS ALEATORIOS

---

Selección de pacientes para el tratamiento experimental
Diferentes criterios de selección de controles no aleatorios y diferentes fechas de acceso
No requerir la obtención del consentimiento informado de los controles
Sesgo de publicación

---

cos o no aleatorios, las comparaciones de tales estudios con ensayos controlados aleatorios de los mismos tratamientos en la misma clase de pacientes han revelado que el resultado experimental es muy similar habitualmente, aunque la diferencia reside en el resultado en el grupo control<sup>1</sup>. Cualquiera que sea el método de comparación de los pacientes seleccionados con los no seleccionados, surgirán diferencias, lo cual es especialmente cierto cuando es preciso solicitar y obtener el consentimiento informado de los pacientes asignados al tratamiento experimental y no de los pacientes control. Puesto que la mayoría de pacientes hace lo que su médico les aconseja y es fácil para el médico influir sobre las de-

---

Publicado en *Med Clin (Barc)* 1992; 98 (8): 305-307.  
Resumen de la ponencia presentada en el simposio sobre "Publicación de ensayos clínicos en revistas médicas" realizado en Madrid (1990) en el seno de la 7ª Conferencia Internacional de Medicina de la Industria Farmacéutica.  
Financiado en parte por la beca n.º 05936-02.

cisiones del paciente, ya sea de modo voluntario o involuntario, el proceso del consentimiento informado ofrece una oportunidad única para diseñar un grupo de pacientes seleccionados sin darse cuenta verdaderamente de ello. Posteriormente, se compara con controles que no han pasado por el proceso del consentimiento informado. Este es un modo de asegurar, voluntaria o involuntariamente, una diferencia.

En estudios sobre el sesgo en la publicación, hay firmes indicaciones de que la inclinación a publicar los resultados positivos y dejar los negativos en el cajón del archivo puede ser más potente en los ensayos clínicos inadecuadamente controlados que en los que sí lo son<sup>2</sup>. El acto de la aleatorización no garantiza que el sesgo se controle adecuadamente en cada ensayo publicado (tabla III).

Hay ejemplos en los cuales se han seleccionado para la aleatorización los pacientes con una previa falta de respuesta al tratamiento control, lo que hace probable que la mitad asignada al tratamiento control no responda tan bien como los pacientes sin el tratamiento previo<sup>3</sup>.

El propósito de la aleatorización consiste en garantizar que los pacientes seleccionados para un ensayo tengan una oportunidad absolutamente idéntica de pertenecer a un grupo terapéutico o a otro. Al admitir pacientes para un ensayo, cualquier sospecha por parte del médico de que sea más probable que un paciente reciba el trata-

TABLA III  
CAUSAS DE SESGO EN EL DISEÑO Y REALIZACIÓN DE LOS ENSAYOS CONTROLADOS ALEATORIOS

Selección de pacientes con antecedentes de mala respuesta a un agente o fármaco control
Proceso de aleatorización no ciego, demostrado por la ausencia o desequilibrio de factores de riesgo prealeatorización (figs. 1-3)
Tratamientos imperfectamente ciegos (demostrado por el conocimiento de los grupos de tratamiento por los investigadores)
Retiradas o abandonos excesivos o desiguales
Frecuente vigilancia de los resultados sin corregir errores alfa o tipo I
Conclusiones inadecuadas a partir de análisis de subgrupos
Elección de objetivos según los datos

miento experimental o el control puede destruir el proceso de aleatorización y distorsionar los resultados. En la figura 1 se ilustra cómo ocurre esto. En la figura 2 se demuestra que esto ocurre; la diferencia de la mortalidad expresada como *odds ratio* de pacientes en ensayos clínicos de diversos agentes terapéuticos para el infarto

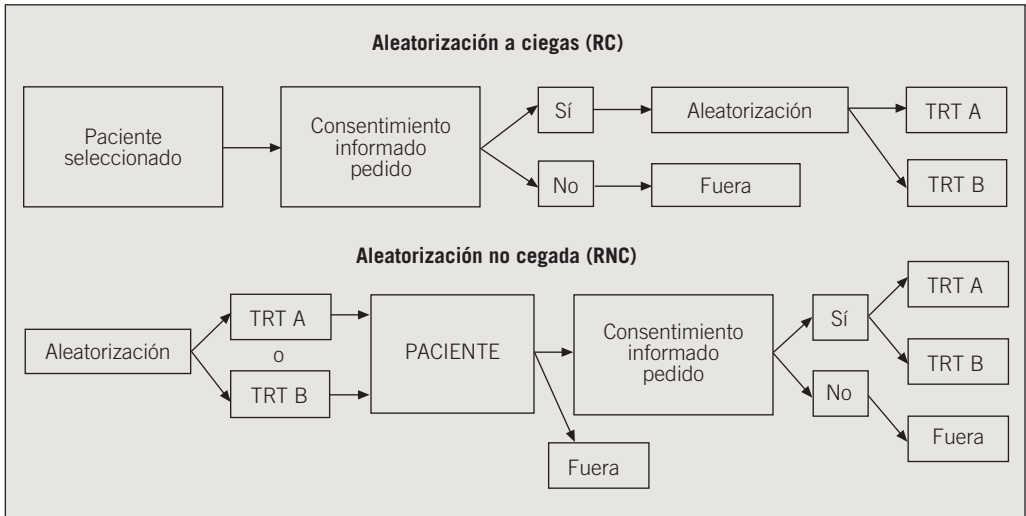


Fig. 1. Esquema de procesos en los que la aleatorización es ciega (RC) y el médico que admite al paciente puede detectar el tratamiento que viene a continuación. TRT: tratamiento.

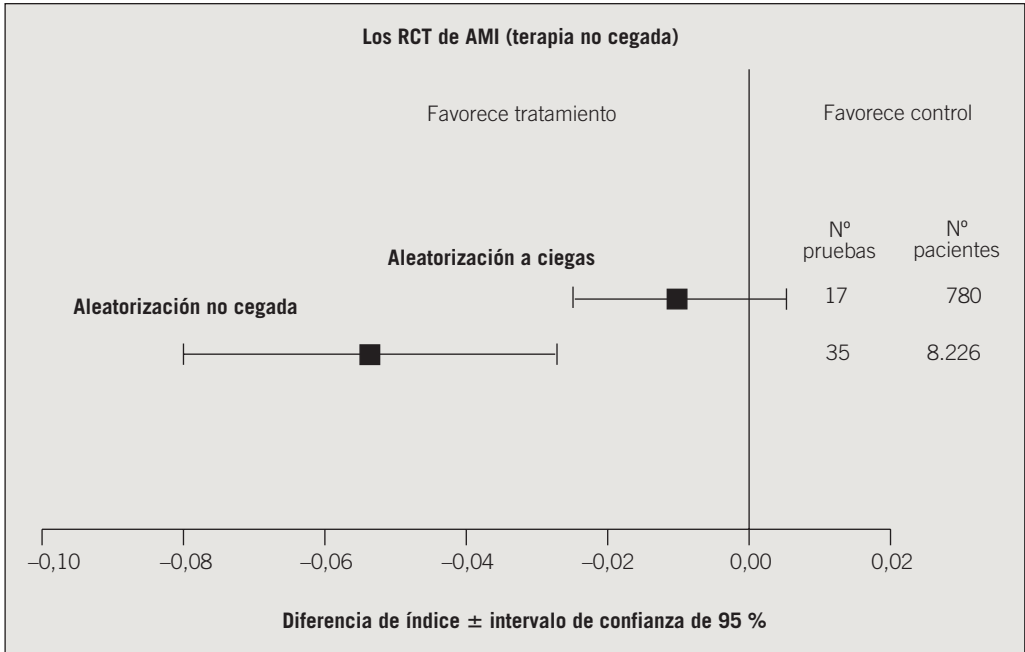


Fig. 2. Resultados de ensayos controlados aleatorios sobre infarto agudo de miocardio, mortalidad como objetivo, en los que el proceso de aleatorización es ciego y no ciego.

agudo de miocardio depende claramente del método de asignación al grupo terapéutico<sup>4</sup>. La diferencia entre el grupo tratado y el grupo control es mínima si la aleatorización es auténticamente ciega de modo que el médico no albergue sospechas sobre qué tratamiento viene a continuación. La diferencia es mayor cuando la asignación se puede determinar mediante la selección de números o fechas de nacimiento, al mirar los sobres a contraluz o al abrirlas con vapor. Los diferentes desenlaces medios podrían explicarse totalmente por el hecho de que los factores de riesgo pretratamiento se distribuyeron al azar en el caso de la aleatorización ciega y se distribuyeron significativamente mal en el caso de la aleatorización no ciega<sup>4</sup>. En otras palabras, cuando el siguiente tratamiento a administrar era conocido por el médico que decidía si admitir o no al paciente, circunstancia seguida por la obtención del consentimiento informado, el grupo tratado, como promedio, presentaba mejores resultados que el grupo control. La importancia de ocultar el proceso de aleatorización (aleatorización ciega) no se ha puesto de relieve tal y como merece en los artículos dedicados al control del sesgo en los

ensayos clínicos. En el mismo grupo de pacientes, la administración u omisión de un placebo no tuvo impacto (fig. 3). La frecuente aparición de una aleatorización inadecuada ha sido bien estructurada por Altman y Dore<sup>5</sup>.

La eficacia de la asignación aleatoria puede ser anulada si se permite que ciertos pacientes abandonen el estudio y otros no. En estudios que no se han realizado auténticamente a doble ciego, es fácil influir sobre el resultado de modo involuntario al mantener en el estudio a los pacientes sometidos al tratamiento experimental con buen resultado y permitir el abandono de los pacientes sometidos a placebo con buen resultado. Por el contrario, se puede permitir que los pacientes tratados con el fármaco experimental con mal resultado abandonen el estudio mientras que se persuade a los del grupo testigo a permanecer en el mismo. Estos actos, habitualmente involuntarios, pueden conducir a una conclusión distorsionada en un ensayo debido al hecho de que las diferencias observadas no obedecen a diferencias en el tratamiento sino en la manipulación de los pacientes<sup>6</sup>. Este problema se reduce por una adecuada ocultación de las terapias, aunque esto

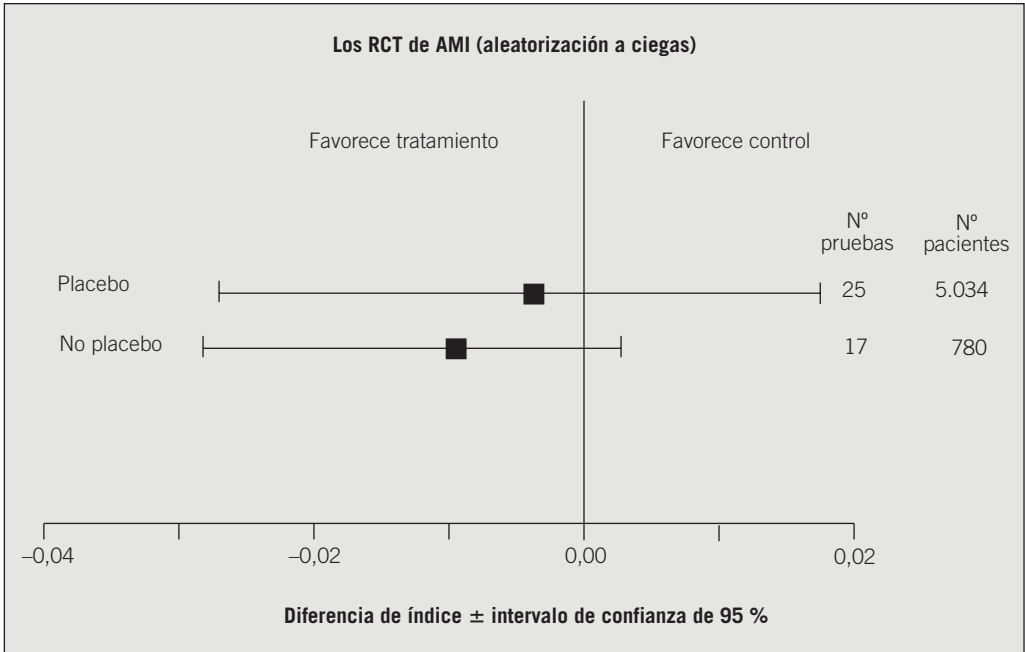


Fig. 3. Resultados de ensayos controlados aleatorios sobre infarto agudo de miocardio, mortalidad como objetivo, cuando se administra placebo o no.

no siempre es posible. El sesgo se reduce cuando todos los pacientes aleatorizados se incluyen en el análisis, aunque puede que no se disponga de la información requerida para ello. Puede suscitarse una legítima sospecha de sesgo cuando el análisis por el tratamiento realmente recibido es diferente del obtenido cuando se analizan todos los pacientes aleatorizados.

Una conclusión sesgada positiva puede ser el resultado de una consulta reiterada de los datos de modo que se termine el estudio tan pronto como la diferencia a favor del grupo tratado descienda a una significación estadística  $p < 0,05$ . La frecuente vigilancia requiere un descenso del valor final de  $p$  si los que observan los resultados pueden actuar sobre ellos. Una conclusión sesgada del estudio puede ser el resultado de lo que se conoce como espolvoreo de datos (*data dredging*) y podría conducir a la notificación errónea de un subgrupo de respuesta como si fuera real sin serlo en el diseño del estudio.

Finalmente, un ensayo correctamente aleatorio ha de tener tan pocos objetivos como sea posible, los cuales se seleccionan en el protocolo y se cumplen en la redacción.

Una vez se ha completado el ensayo, hay muchas oportunidades para el sesgo en el proceso de redacción (tabla IV).

El sesgo en la publicación ha sido bien estudiado<sup>7-9</sup> y puede suceder en todos los niveles desde el autor hasta la aceptación o rechazo finales por los directores de las revistas. De hecho, parece que se relaciona más frecuentemente con las decisiones de los autores que con el rechazo de los estudios negativos por parte de los

TABLA IV  
CAUSAS DE SESGO EN LA PUBLICACIÓN DE ENSAYOS CONTROLADOS ALEATORIOS

Publicación de sesgo
Publicación repetida del mismo material
Selección sesgada de referencias
Inconsistencias de datos en resumen, texto, discusión y tablas
Ausencia de declaración de posibles conflictos de intereses
Ausencia de revisión adecuada



directores, aunque esto último ciertamente ocurre<sup>8</sup>. La publicación repetida del mismo material cuando el efecto terapéutico es impresionante se puede considerar como una forma de publicación sesgada de los ensayos clínicos<sup>10</sup>. La selección sesgada de referencias que respaldan las conclusiones del ensayo publicado se ha demostrado claramente en la literatura de los ensayos clínicos<sup>11</sup>. A veces hay cifras diferentes en el resumen, texto, tablas o discusión y, si se asume que éstos son errores involuntarios, se pueden considerar como oportunidades para el sesgo. La falta de una declaración categórica de posibles conflictos de intereses por parte de los investigadores puede suscitar una legítima sospecha de que pueda haber, de hecho, algún conflicto de intereses.

Finalmente, uno de los modos más eficaces de reducir el sesgo en la publicación de los ensayos clínicos es el requerimiento de un adecuado mecanismo de revisión para la publicación de todos ellos. Al menos, dos revisores deberían evaluar críticamente todos los ensayos. Personalmente, creo que se obtiene un resultado más eficaz y honesto si los revisores están dispuestos a firmar al pie de sus revisiones, aunque entiendo que muchas personas aún creen en la importancia del anonimato en el proceso de revisión. De todos modos, los ensayos clínicos no deberían publicarse jamás sin haber pasado por el proceso de la revisión. La tabulación sobre los diferentes modos en que se puede sesgar la publicación de ensayos clínicos es el resultado de una gran cantidad de experiencia en la revisión y análisis de los mismos. No es una revisión exhaustiva de la calidad de los ensayos; sin embargo, incluye los te-

mas más importantes y si los investigadores son conscientes de ello, habrá una reducción máxima del sesgo en las publicaciones.

## BIBLIOGRAFÍA

1. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72: 233-240.
2. Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989; 84: 381-392.
3. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal, antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989; 10: 31-56.
4. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309: 1.358-1.361.
5. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-153.
6. Bhaskar R, Reitman D, Sacks HS, Smith H, Chalmers TC. Loss of patients in clinical trials that measure long-term survival following myocardial infarction. *Controlled Clin Trials* 1986; 7: 134-148.
7. Sterling T. Publication decisions and their possible effect on inferences drawn from test of significance or viceversa. *J Am Stat Assoc* 1959; 54: 30-34.
8. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. *Controlled Clin Trials* 1987; 8: 343-353.
9. Rosenthal R. The file draw problem and tolerance for null results. *Psychol Bull* 1979; 86: 638-641.
10. Gotzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989; 36: 429-432.
11. Gotzsche PC. Reference bias in reports of drug trials. *Br Med J* 1987; 295: 654-656.

---

# La ética de la publicación: el caso de los resultados negativos

---

Sergio Erill

Laboratorio Dr. Esteve. Barcelona.

Es difícil concebir cualquier otra respuesta que no sea un sí incondicional a la pregunta de si los estudios en los que se obtienen resultados negativos deberían ser publicados o no. Cualquier estudio bien diseñado y bien realizado es una contribución al conocimiento, independientemente de la magnitud, dirección o significación estadística de los resultados obtenidos, y la comunicación es algo inmanente a la ciencia. Sin embargo, uno se siente tentado a restringir la pregunta, puesto que hay más de un tipo de resultado negativo<sup>1</sup>. Así, pues, es posible considerar que los resultados de un estudio que son opuestos a la dirección de la hipótesis de prueba son negativos aun cuando sean estadísticamente significativos mientras, con bastante frecuencia, el término se aplica a los resultados que no alcanzan el nivel preestablecido de significación estadística. Por si esto no produjera ya suficiente confusión, no es raro oír hablar de estudios negativos y, aunque algunos han discutido acaloradamente contra la utilización de este término tan impreciso<sup>2</sup>, continúan apareciendo, incluso en los títulos de los editoriales de revistas de primera categoría<sup>3</sup>.

En cualquier circunstancia, parece claro que debería existir un lugar para los resultados negativos, y esto no solamente atañe a los ensayos clínicos, sino también a todo estudio fundamentado. Se hace referencia frecuentemente<sup>4-6</sup> a los archivos de la industria farmacéutica en cuanto a que contienen datos que no se han publicado para evitar cualquier desdoro de la imagen del compuesto al cual se refieren o del laboratorio si el producto nunca se llegara a comercializar. Teniendo en consideración la naturaleza y los propósitos de los estudios de toxicidad que se emprenden actual-

mente y la bien conocida tasa de desgaste de las nuevas entidades químicas en el proceso de descubrimiento y desarrollo farmacológico, resulta bastante claro que se perdería muy poca imagen si estos datos se publicaran. Sin embargo, se puede construir un caso para la necesidad de publicar datos, tanto sean positivos como negativos resultantes de estudios preclínicos. Así pues, la disponibilidad de datos acerca de la actividad antimicrobiana de una multitud de compuestos sulfamídicos (algunos de ellos muy activos pero bastante pobres como antimicrobianos) permitió establecer relaciones estructura-actividad<sup>7</sup> que condujeron a decisiones racionales acerca del desarrollo de nuevas sulfamidas. Este éxito debe compararse con la actual imposibilidad de discernir cuáles son los rasgos que conducen a la mutagenicidad en la serie de quinolonas antibacterianas, en la que algunos compuestos están absolutamente exentos de este efecto mientras otros tienen una potente actividad. Es imposible saber si la disponibilidad de datos sobre mutagenicidad, ya generados pero nunca publicados, de un sustancial tanto por ciento de los cientos de quinolonas sintetizadas hasta la fecha, ofrecería claves definitivas para los determinantes de mutagenicidad en esta serie, aunque está bastante claro que ello favorecería el progreso científico.

Por otro lado, en vista de las dificultades universalmente reconocidas en lo que concierne a la extrapolación de los datos sobre toxicidad animal al hombre, resulta obvio que la publicación de los resultados de estudios en animales y seres humanos con fármacos, en los que se ha observado que son demasiado tóxicos como para ser comercializados, fomentaría indudablemente el conocimiento científico. Parece casi inevitable que los que hacen una profesión de la crítica de la industria farmacéutica se sentirían muy tentados a utilizar esta clase de datos para sus objetivos particulares, aunque quizá debiéramos citar a Modell<sup>4</sup>, quien, al solicitar información sobre fra-

---

Publicado en *Med Clin (Barc)* 1992; 98 (8): 308-309.  
Resumen de la ponencia presentada en el simposio sobre "Publicación de ensayos clínicos en revistas médicas" realizado en Madrid (1990) en el seno de la 7ª Conferencia Internacional de Medicina de la Industria Farmacéutica.

casos y defectos farmacológicos a los laboratorios, nos recordó que, en interés de la difusión de la Palabra, incluso la Biblia revelaba detalles sobre el mal. Si nos centramos en los resultados de los ensayos clínicos, tampoco hay duda que los resultados negativos son de capital interés si surgen de estudios bien diseñados y bien realizados cuyo objetivo sea comprobar una hipótesis fundamentada. Poco se puede decir sobre la importancia del diseño y realización de un ensayo clínico, aunque el papel que desempeña la naturaleza de la hipótesis de trabajo por lo que respecta a la conveniencia de la publicación de los resultados de un ensayo puede resultar algo confuso. Para clarificar este punto, es importante mencionar un ejemplo sugerido por Angell<sup>9</sup>, la directora ejecutiva del *New England Journal of Medicine*, en el cual hace referencia al valor muy limitado de los resultados de un ensayo en el que se demostraba que la penicilina no era más eficaz que el placebo frente al cáncer, sin importar cuán cuidadosamente diseñado y realizado haya sido el ensayo. Al considerar las razones en favor de la publicación de los ensayos en los que se obtienen resultados negativos, probablemente lo más sencillo sea tratar, en primer lugar, del caso de los ensayos cuyos resultados tienen una tendencia, pero no alcanzan significación estadística. Una referencia a la actualmente frecuente técnica del metaanálisis probablemente bastaría para sustentar la idea, aunque merece la pena recordar que la significación estadística de las diferencias observadas sólo es uno de los elementos de interés de un ensayo clínico. El conocimiento de los efectos de un fármaco sobre el ser humano no puede ser restringido dentro de las fronteras de un ensayo clínico o más, utilizando una dosis determinada durante un tiempo determinado. No es posible prescribir un fármaco con conocimiento de los efectos temporales implicados en la respuesta o con la naturaleza y configuración de la relación dosis-respuesta. Además, resulta bastante claro que, en algunas condiciones patológicas, hay pacientes simplemente refractarios o fármacos perfectamente válidos. El conocimiento comprensivo de los efectos de un fármaco en el ser humano sólo puede surgir a partir de multitud de estudios en los cuales los datos, y no sólo la dirección o la significación estadística de los resultados, sean lo importante finalmente. Todas estas razones se podrían aplicar fácilmente a los ensayos clínicos en los cuales no se detecta una diferencia estadísticamente significativa o incluso una tendencia a la misma en el análisis de los resultados, y aun cuando se asumiera que el diseño y tamaño del ensayo son lo suficientemente convincentes como

para que la observada falta de diferencia reflejara verdaderamente una equivalencia, hay muchas razones a favor de la publicación de dicho ensayo. Podríamos considerar, por ejemplo, la necesidad de comparar estos resultados con los de otros ensayos en los que se indique lo contrario, o incluso la necesidad de evitar la exposición de más pacientes a una intervención similar. Se podría imaginar, por ejemplo, el caso de un investigador clínico o un laboratorio comercial que tiene una idea brillante sobre una posible utilización de un determinado inhibidor de la ciclooxigenasa. Se diseña cuidadosamente un estudio que se lleva a cabo y, desgraciadamente, se demuestra definitivamente que el fármaco no es más eficaz que el placebo, a pesar de la solidez de la hipótesis. Si este estudio no se publicara, y el presunto efecto dependiera de la inhibición de la ciclooxigenasa, es bastante probable que la misma idea se le ocurriera a otros investigadores o laboratorios comerciales que trabajaran con el mismo fármaco u otro inhibidor de la ciclooxigenasa. Se planearía un nuevo estudio y otros pacientes serían expuestos a un fármaco que, después de todo, demostraría carecer de valor y que, como cualquier otro, conlleva cierto riesgo.

Así pues, en interés de la ciencia y de la sociedad, los estudios clínicos en los que se obtienen resultados negativos se deberían publicar. Varios estudios empíricos y una gran cantidad de pruebas anecdóticas indican que esta situación dista mucho de ser la ideal<sup>8-11</sup> y se ha señalado un número de recomendaciones, incluida la necesidad de creación de registros internacionales de ensayos clínicos<sup>12,13</sup>. Los sistemas para disminuir los sesgos en la publicación de ensayos clínicos se comentarán posteriormente en este simposio, aunque merece la pena mencionar aquí que *Clinical Pharmacology and Therapeutics*, una revista puntera en el campo, publicó un editorial en 1973 en el que abogaba por este tipo de datos<sup>4</sup>, y que otra revista venerable dedicada a la misma disciplina, la *Journal of Clinical Pharmacology* publicó un editorial parecido hace algunos meses<sup>5</sup>. Es interesante que, en el momento en el que *Clinical Pharmacology and Therapeutics* demandaba intensamente la publicación de datos clínicos sobre fármacos que nunca alcanzarán el mercado, el tanto por ciento de publicaciones de ensayos clínicos en sus páginas decreció rápidamente (en favor de los datos de farmacocinética). Por lo que concierne al *Journal of Clinical Pharmacology*, éste ha incluido durante muchos años, entre las noticias para los autores, una solicitud de artículos que presentaran resultados negativos, pero raramente ha publicado alguno.

No sería adecuado dar recomendaciones, aunque, como quiera que estamos tratando con dos revistas pioneras en el campo de la farmacología clínica, recordemos lo que el editor actual de *Clinical Pharmacology and Therapeutics* ha escrito recientemente<sup>14</sup> en relación a la práctica, demasiado frecuente actualmente, que consiste en hinchar el *curriculum vitae*: ha llegado el momento de detener la actual práctica consistente en evaluar a las personas por el recuento de sus publicaciones en lugar de leer y evaluar sus trabajos publicados. Toda petición de publicación de estudios en los que se obtengan datos negativos puede parecer un aliado a la aparición y éxito de revistas poco rigurosas. La ciencia necesita datos, cualquiera que sea su signo, siempre y cuando se deriven de estudios cuidadosos. Es una falta de ética privar a la comunidad científica de los datos negativos simplemente porque son negativos, pero sería igualmente una falta de ética ser permisivo con respecto a la manera en que se generan. Toda solicitud para la publicación de los resultados negativos nunca debería basarse en la benevolencia sobre cualquier tipo de trabajo en detrimento de su calidad.

## BIBLIOGRAFÍA

1. Dickersin K. Confusion about "negative" studies. *N Engl J Med* 1990; 322: 1.084.
2. Chalmers I. Proposal to outlaw the term "negative trial". *Br Med J* 1985; 290: 1.002.
3. Angell M. Negative studies. *N Engl J Med* 1989; 321: 464-466.
4. Modell W. Failed and flawed drugs: a frozen resource. *Clin Pharmacol Ther* 1973; 14: 771-772.
5. Somberg JC. My data, your data, whose data? *J Clin Pharmacol* 1990; 30: 1-2.
6. Hemminki E. Study of information submitted by drug companies to licensing authorities. *Br Med J* 1980; 280: 833-836.
7. Bell PH, Roblin ROJR. Studies on chemotherapy. VII: a theory of the relation of structure to activity of sulfanilamide type compounds. *J Am Chem Soc* 1942; 64: 2.905-2.917.
8. Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979; 86: 638-641.
9. Dickersin R, Chan S, Chalmers TC et al. Publication bias and clinical trials. *Controlled Clin Trials* 1987; 8: 843-853.
10. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989; 84: 374-380.
11. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1.385-1.389.
12. Maxwell C. Clinical trials, reviews, and the journal of negative results. *Br J Clin Pharmacol* 1981; 11: 15-18.
13. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986; 4: 1.529-1.541.
14. Reidenberg MM. Padding the curriculum vitae: impairing the utility and integrity of the biomedical literature. *Clin Pharmacol Ther* 1989; 45: 337-339.

---

# Editoriales, revisiones y cartas: por quién, para quién y cuándo

---

Philip Brown

Director de Scrip. Richmond. Reino Unido.

El tiempo no me permite comentar con detalle este importante tema. En 10 minutos tan sólo, únicamente puedo tocar los principios que atañen a estas decisiones y dejar la exploración complementaria de estos temas y aspectos específicos a la discusión que sigue a mi presentación.

He omitido también las palabras ensayo clínico del título de mi presentación. Creo que, si nos referimos a los ensayos clínicos, las decisiones concernientes a los mismos están basadas en los mismos principios. No quisiera que se crea que mis puntualizaciones se refieren únicamente a los informes sobre ensayos clínicos.

A modo de introducción, podría decir que los contenidos de los editoriales, revisiones y cartas al director y las decisiones editoriales sobre la publicación de comentarios y descripciones constituyen una fuente de fricción interminable entre todas las partes implicadas: el autor del material, el director y aquellos que consideran que ellos mismos o la causa médica que apoyan, han sido injustamente atacados.

En Reino Unido, por ejemplo, los representantes de la industria farmacéutica, han manifestado en repetidas ocasiones su preocupación a los directores de *The Lancet* y *British Medical Journal* sobre la publicación de informes aislados de reacciones adversas a productos farmacológicos en la sección de cartas al director y han pretendido que se les tuviera en cuenta para las decisiones sobre cómo publicar dicha correspondencia y si era conveniente o no. En el debate en curso sobre los posibles efectos adversos de diversos elementos de nuestra dieta (hidratos de carbono, azúcar, productos de la industria Táctica), se han suscitado fuertes sentimientos por las posturas antagónicas adoptadas desde los Comités editoriales de las revistas médicas.

Los directores de las revistas de circulación controlada, que son aquellas publicaciones que dependen de la publicidad comercial o del patrocinio para su viabilidad financiera, a diferencia de las revistas cuya fuente de ingresos es la suscripción, reciben actualmente amenazas implícitas o explícitas de sus anunciantes para que no publiquen artículos o editoriales hostiles. No es inusual que las compañías farmacéuticas manifiesten su protesta, si aquello sucediera, en forma de retirada de su apoyo financiero a través de los anuncios comerciales.

Naturalmente, no siempre hay víctimas de los editoriales, revisiones y cartas al director, aunque hay el número suficiente como para que esta presentación sea un componente necesario de este simposio satélite.

Como anterior director del *Scrip World Pharmaceutical News* y actualmente editor de esta revista y de otras cuatro de la compañía PJB Publications, trataré este tema desde la posición del director convertido en editor. Debería añadir que, como quiera que las publicaciones de la compañía a la que pertenezco están financiadas a través de suscripciones y son esencialmente revistas de novedades, no estamos expuestos con tanta frecuencia a la controversia de la que estoy tratando aquí, aunque no desconocemos que estamos sometidos a presión por parte de las compañías implicadas en controversias para que omitamos o rectifiquemos un artículo conteniendo alguna novedad.

La piedra de toque para tratar de este tema reside en el reconocimiento de un hecho: en la publicación de revistas médicas o, por cuanto a este punto atañe, en cualquier modalidad de publicación, el director es rey. El director establece todos los criterios para la publicación (la política editorial, el perfil de lectores y el contenido de los editoriales). Para ser un director efectivo, debe estar dotado de libertad en estas tres áreas y, para ser un buen director, debe ser lo suficientemente firme como para mantenerse en su posición en todo momento y bajo todas las formas de presión. Al igual que sucede con los presidentes, en el mundo de la publicación, el director asume toda la responsabilidad.

---

Publicado en *Med Clin (Barc)* 1992; 98 (8): 309-310.  
Resumen de la ponencia presentada en el simposio sobre "Publicación de ensayos clínicos en revistas médicas" realizado en Madrid (1990) en el seno de la 7ª Conferencia Internacional de Medicina de la Industria Farmacéutica.

Exploremos juntos un poco más los criterios clave de la política editorial, el perfil de lectores y el contenido de los editoriales, puesto que se relacionan directamente con el tema de esta presentación.

Adjudicado este papel soberano al director, se desprende que sus decisiones sobre qué publicar, por quién, para quién y cuándo, se convierten en la política editorial de la publicación de la cual él es responsable. Como cada director es un individuo, cada publicación tiene su propia política editorial individual.

Pero, dado que la mayoría de revistas científicas o médicas están consideradas como proveedoras de información sumamente objetiva, ¿es correcto decir que dichas revistas tienen políticas editoriales? ¿Ejerce el director del *New England Journal of Medicine* el mismo poder decisorio en política editorial que sus colegas, por ejemplo, en el mundo de los periódicos o revistas no biomédicas?

La respuesta es sí. El sesgo editorial introducido al seguir la política editorial es tan significativo en *New England Journal of Medicine* como, por ejemplo, en *The Times*, *Frankfurter Allgemeine* o *El País*. La información, ya sea por su selección o por el modo en que está escrita, está siempre sesgada según las consideraciones de política editorial. El lector obtiene lo que le dan. De modo esperanzador, ésta es la verdad, aunque, raramente, si es que alguna vez sucede, toda la verdad.

Desde el punto de vista de los lectores, si el contenido editorial concuerda con sus opiniones y percepciones, la publicación es *de facto* objetiva para ellos. Si no concuerda, los lectores encuentran otra publicación que sí lo hace. Por lo tanto, si los lectores van generalmente a una con el contenido editorial de una publicación, el director y los lectores se encuentran en la misma longitud de onda de opinión, comparten las mismas opiniones y están sujetos a los mismos sesgos. Utilizando la analogía del espejo, una publicación tiene éxito con sus lectores si refleja exactamente sus opiniones del mundo, de la vida, del universo y de todo.

De lo que acabo de decir, se desprende que un director debe definir a su público para su revista. Debe saber para quién escribe, a quién ofenderá y a quién complacerá. Pero ningún director cumple simplemente con complacer a sus lectores, también debe estimular y entretener. Para lograrlo, orquestará la información para mantener a su leal y definido público. Les dará las buenas y las malas noticias, suscitará pasiones, estimulará los ánimos. Mantener a los lectores intelectual y emocionalmente en vilo es un ingrediente esencial de la táctica de juego del director para conseguir el éxito.

Se puede pensar que los lectores de revistas médicas componen un grupo bastante uniforme

de personas y que, por tanto, las consideraciones hacia el público lector no son tan importantes como cuando se trata de dirigir una publicación para diversos sectores del público general. Yo señalaría, sin embargo, que dentro de la profesión médica hay escuelas de pensamiento bastante diferentes en lo tocante al modo en que la medicina debería practicarse científica, económica y moralmente. El director de una revista médica debe escoger entre varias opciones en cuanto a las líneas que seguirá sobre las controversias del día y su elección se relacionará con los grupos de la profesión que suscriban dichas opiniones.

Una vez ya presentado el tema de la política editorial y de la selección del público lector, me referiré al contenido editorial. El director ejercita su política editorial mediante la selección de temas y la presentación editorial. También tiene la responsabilidad de informar y satisfacer las necesidades informativas de sus lectores. Sobre este punto, quisiera insistir en la importancia de la información como el ingrediente esencial para el cambio, desarrollo y mejoramiento. Si la medicina debe progresar, los mercados deben desarrollarse o el pensamiento político debe progresar, debe existir un flujo libre y total de todas las formas de información (que cubra los hechos, opiniones y análisis).

Todo director de una publicación que presente noticias, comentarios y análisis serios, tiene la responsabilidad de ser tan objetivo como sea posible sobre el contenido informativo de su publicación. Debe resistir la presión para publicar por otros motivos que no sean noticias de interés público o que provengan de opiniones autorizadas. En el momento en que una publicación deviene infiltrada por la propaganda o contaminada por el apoyo financiero de organizaciones o individuos con intereses concretos, pierde su credibilidad y su *status*. Puede sobrevivir, pero será una sombra de lo que fue.

Ante este breve esquema de los principios esenciales de la publicación, es obvio que la respuesta a las preguntas formuladas en el título de mi presentación, "Editoriales, revisiones y cartas: para quién, por quién y cuándo", es simplemente lo que al director incumbe decidir y, por lo tanto, asumir la responsabilidad de sus decisiones. Obviamente, un director debe ser consciente, debe buscar asesoramiento y debe consultar para tomar decisiones que sean defendibles. Ningún director debe operar en el vacío.

Un director también aceptará implícitamente el derecho de réplica y será tan tenaz en la defensa de este principio como lo debe ser en la presentación de su propia síntesis sobre la verdad actual a través de las páginas de su publicación.

Creo que incumbe al director decidir qué editoriales y revisiones encargar y a quién. Además, el director debe decidir qué cartas publicar. También debe decidir cuándo las publicará. Si está dotado de una actitud luchadora, las publicará al principio de la controversia en curso y si no, se satisfará en la revisión de los hechos pasados y en extraer enseñanzas de los mismos. Diría que he notado que las revistas médicas están cayendo en la premura generalizada de publicar tan pronto como sea posible. Existe un creciente elemento novedoso en lo que aparece publicado junto con una pérdida de peso específico, lo cual, posiblemente, lamentemos. Volviendo a lo que expuse al principio, la industria farmacéutica se ha preocupado por los informes aislados, sobre reacciones adversas a medicamentos publicados en revistas principales. El motivo de esta preocupación surge del impacto de dichas cartas sobre el público lector, lo cual plantea otra vez la importante pregunta: ¿quiénes son los lectores de las revistas médicas?

En esta época ilustrada, es muy difícil restringir cualquier información publicada únicamente al público inmediato al que fue dirigida. Tarde o temprano, toda la información significativa publicada alcanza el dominio público y será conocida por todo el mundo.

Dada la intensidad del apetito público por las noticias médicas, no es sorprendente que el informe médico de esta semana sea titular de periódico la semana que viene, si no lo es mañana. Efectivamente, en algunos países, los editores de revistas médicas alientan la publicación periodística mediante el envío de copias prepublicación con la única restricción de la fecha de llegada al buzón de los lectores.

Volvamos a la preocupación de la industria farmacéutica sobre la publicación de reacciones adversas esporádicas en las cartas al director. El público no sólo consiste en la profesión médica, sino también en el público general. Y la información no se presenta como una sobria carta enterrada en una sección de 20 páginas junto con otras 40 cartas, sino posiblemente bajo unos titulares a toda plana en lenguaje simplificado que no puede reflejar las advertencias, implícitas o explícitas, de la contribución general. Existe la dura evidencia de que los informes sobre reacciones adversas medicamentosas aisladas trasladados a periódicos sensacionalistas han disparado la preocupación masiva del público, la intervención gubernamental reguladora y la retirada de productos.

Los problemas del tema del amplio público lector también tienen relación con las revisiones y editoriales más controvertidas que se publican en las revistas médicas y plantean importantes pregun-

tas sobre el modo cómo un director define su política editorial y a su público, al cual se dirige dicha política. ¿Debe aceptar el director, como un criterio importante, el hecho de que no produce simplemente una publicación médica para un público médico cultivado? ¿Debe aceptar que sus consideraciones políticas incluyen el hecho de que otros directores y periodistas menos informados de otras publicaciones manipularán este material y que, detrás de ellos, hay millones de lectores no englobados en las definiciones de público lector?

Creo que la respuesta básica es que el director tiene un deber con su público primario y no con los estratos secundario y terciario del público lector. El motivo para esto es que la responsabilidad del director es difundir la información y las opiniones a quien las necesita. Tal como he dicho, la información contenida en las publicaciones cumple la función de originar el cambio y la modificación de opiniones y actitudes.

Todo lo que desprecie el valor de esta información o de su claridad anula el objetivo básico de su publicación.

Yo considero que los efectos de la publicación en revistas médicas sobre audiencias más amplias son responsabilidad de otros directores. Del mismo modo que el director del *British Medical Journal* es responsable del público lector de dicha publicación, el director del *New York Times* es el responsable del suyo y el director del programa televisivo estadounidense *60 minutes* del suyo y así sucesivamente.

Acepto que puede haber errores en esta política, pero entonces, hay que preguntarse quién es el responsable de los errores.

Si la confusión y la alarma son el resultado de la presentación en los medios de comunicación de la información extraída de las revistas médicas, la responsabilidad de esta contingencia recae sobre los periodistas y directores de los medios de comunicación. Es su responsabilidad ser conscientes de los riesgos que corren cuando publican información médica y garantizar que ésta se presente correctamente. Sería completamente erróneo decir que la información no debería ser publicada en la revista médica primaria porque se presentará de modo distorsionado por subsiguientes directores.

Espero que, aun cuando no haya tratado de temas específicos, la respuesta a las preguntas formuladas en el título, sea más sencilla. En estos temas, la persona que toma las decisiones es el director. La decisión sobre qué debería publicarse y cuándo nunca la tomarán personas ajenas al director. Sin embargo, hay un límite: si a los lectores o a los patrocinadores no les gusta lo que se ha escrito, siempre pueden retirar su apoyo.

Parte 8

---

**REVISIONES BIBLIOGRÁFICAS**

---



---

# La medicina basada en pruebas: revisiones sistemáticas. La Colaboración Cochrane

---

Jesús María López Arrieta<sup>a</sup> y Nawab Qizilbash<sup>b</sup>

<sup>a</sup>Médico adjunto. Servicio de Geriátría. Coordinador de la Unidad de Memoria. Hospital de Cantoblanco. Servicio Regional de Salud. Madrid. Visiting Fellow. Memory Trials Research Group.

<sup>b</sup>Senior Lecturer. Memory Trials Research Group. Clinical Geratology Division. Radcliffe Infirmary. Oxford. Reino Unido. Co-Ordinating Editor. Cochrane Dementia and Cognitive Impairment Group. The Cochrane Collaboration. Oxford. Reino Unido.

---

## Archie Cochrane

Archie Cochrane fue un epidemiólogo británico que desde sus comienzos profesionales tuvo una gran sensibilidad por la justicia distributiva. Luchador infatigable a favor de las clases sociales menos favorecidas, su primitivo ideal giraba en torno a un sistema de salud libre, gratuito y universal. Sus convicciones políticas le llevaron a España, donde participó en actividades antifascistas y posteriormente participó en el frente europeo, durante la Segunda Guerra Mundial. Permaneció 4 años prisionero por los alemanes, y ejerciendo como médico en un campo de concentración, tuvo ocasión de conocer la fortaleza humana. A pesar de encontrarse en condiciones infrahumanas, hacinado en un campo de concentración con 20.000 prisioneros de guerra, con una dieta de 600 calorías diarias, padeciendo todo tipo de enfermedades y sin medicamentos, sólo 4 cautivos murieron, tres de ellos por heridas de arma. Si hubiese dispuesto de suficientes medios terapéuticos, el joven Cochrane los habría considerado determinantes del escaso número de bajas; es decir, hubiera asociado erróneamente el curso habitual de la enfermedad hacia la curación con la intervención terapéutica. Esta falsa asociación se ha denominado la mentira de la regresión (*regression fallacy*)<sup>1</sup>, en clara referencia al concepto estadístico de regresión a la media de Galton y Pearson<sup>2</sup>.

Después de la guerra, Cochrane fue testigo del nacimiento y desarrollo del Servicio Nacional de Salud (SNS) británico. Un auténtico hito social, donde todas las personas, ricas o pobres, tendrían la garantía de recibir la misma calidad de atención sanitaria, sin escatimar medios para ello. No tardó en percatarse de que este sistema

de salud era costoso e ineficaz. Durante los primeros 15 años del SNS, los gastos se incrementaron desenfadadamente, las peticiones de pruebas complementarias se multiplicaron por tres y los servicios de radiodiagnóstico se duplicaron. El SNS disponía de un cheque en blanco con el que satisfacer las demandas de los pacientes y los deseos de los médicos de ayudar: utilizando nuevas pruebas diagnósticas (sutilmente animados por los investigadores) y probando medicamentos novedosos (hábilmente persuadidos por las compañías farmacéuticas). Cualquier organización industrial de un tamaño similar al SNS hubiese dispuesto de una sección de investigación para evaluar la efectividad de sus servicios. Desafortunadamente, por aquel entonces no existía ningún método para estudiar los efectos de la atención sanitaria. Pero no hubo que esperar mucho tiempo pues en 1948, de la mano de Sir Austin Bradford-Hill apareció una técnica que revolucionaría la investigación clínica, el ensayo clínico aleatorizado (ECA). Éste es un experimento por el que se compara, al tiempo, una intervención terapéutica, diagnóstica o preventiva con otra intervención, con placebo o con ninguna intervención, para conocer su eficacia y seguridad<sup>3</sup>. Ese mismo año Cochrane comenzó a participar en ensayos clínicos dentro de la neumología, llegó a ser director de una Unidad de Epidemiología del Consejo Británico de Investigación Médica y fue el primer decano de la Facultad Británica de Medicina Comunitaria. Su cuerpo de pensamiento quedó plasmado en su libro *Effectiveness and Efficiency: Random Reflections on Health Services* publicado en 1972. En él recalca que los recursos sanitarios son limitados, por lo tanto, no deberían malgastarse, y sólo habría que aconsejar aquellos tratamientos cuya eficacia hubiese sido demostrada de forma clara. Considera que la toma de decisiones en la asistencia sanitaria debe basarse en los ECA<sup>4</sup> y juzga igual de im-

---

Publicado en *Med Clin (Barc)* 1996; 107 (15): 581-585. Trabajo financiado parcialmente con la ayuda FIS/0477.

portante la revisión sistemática (RS) de éstos como la divulgación de sus resultados. Para ello propuso la creación de una red internacional de revisores que colaborasen en la elaboración, actualización y difusión de RS de ECA sobre los efectos de la atención sanitaria<sup>5</sup>. En 1979 Cochrane censuró a la profesión médica por no haber organizado un resumen crítico por especialidades y revisado periódicamente de todos los ECA sobresalientes<sup>6</sup>.

### **El reto de Archie Cochrane**

Las ideas de Cochrane fueron progresivamente aceptadas por la comunidad médica, pero la aplicación práctica de sus principios resultaba utópica, principalmente por la dificultad de acceso a la información relevante, aunque ésta se encuentre publicada en libros o revistas<sup>7</sup>. Los ECA publicados como artículos originales son muy numerosos y se encuentran dispersos por la inabarcable bibliografía médica, lo que los convierte en poco útiles, al llegar sólo a un pequeño tanto por ciento de los profesionales de la salud. Por ello, la mayoría de los médicos confían en las revisiones de los artículos originales, como una forma de poder enfrentarse a esta ingente información<sup>8</sup>. Desafortunadamente en la mayoría de los casos se trata de revisiones narrativas que no cumplen los principios de la investigación original, especialmente el control de los errores sistemáticos y aleatorios<sup>9</sup>.

Durante el siguiente decenio hubo investigadores, especialmente los que habían participado en ECA sobre terapia oncológica, sobre antiagregación para la prevención secundaria en las enfermedades cardiovasculares y sobre descontinuar la antibioterapia del tubo digestivo en pacientes sometidos a cuidados intensivos, que comenzaron a atender las críticas de Cochrane. Continuaron su labor investigadora, realizando RS<sup>10-13</sup> en sus respectivos campos. Estas primeras tentativas de colaboración multicéntrica han aportado información terapéutica trascendental y han establecido directrices para la investigación futura y han servido como referencias de calidad para juzgar RS de ECA<sup>14</sup>.

Cochrane vivió lo suficiente para contemplar cómo sus enseñanzas habían comenzado a calar en la comunidad médica. En 1987, el año anterior a su muerte, se refirió a la RS de ECA sobre obstetricia realizados por el NPEU (National Perinatal Epidemiology Unit) como un hito en la historia de los ECA y en la evaluación de los cuidados y aconsejó a las demás especialidades que copiaran sus métodos<sup>15</sup>.

### **Revisión de la bibliografía. Revisión sistemática**

Cada vez es mayor el número de ECA que se publican, y es difícil encontrar una información resumida de los estudios más relevantes sobre un tratamiento determinado. Dos millones de artículos salen a la luz anualmente en la bibliografía biomédica, repartidos en más de 20.000 revistas<sup>16</sup>. La RS de esta bibliografía científica es necesaria para depurar toda esta desmesurada cantidad de información<sup>17</sup>. Para poder hacer frente a toda esta divulgación, la mayoría de los médicos confían en el conocimiento personal de unos pocos ECA representativos o bien hechan mano de los trabajos de revisión clásica en revistas o en libros de texto. Desgraciadamente la calidad de la mayoría de las revisiones médicas dejan mucho que desear, debido a que las labores de recogida, análisis y publicación de los resultados de una revisión no se realizan con el mismo cuidado y detalle que cuando se trata de un trabajo de investigación original. Esta falta de método tiende a extraer conclusiones que, en el mejor de los casos, están sesgadas, y en el peor de los casos son erróneas<sup>6,18</sup>.

Antman et al comprobaron que los principales libros en inglés de cardiología mostraban algunas indicaciones terapéuticas en el infarto agudo de miocardio (IAM) con un desfase de más de 10 años. Por ejemplo, el tratamiento trombolítico no lo recomiendan la mayoría de estos prestigiosos tratados hasta finales de los años ochenta, en cambio ya había evidencia de que prolongaba la supervivencia de los pacientes con IAM desde antes de 1975, y lo que es peor, verificaron cómo, todavía en 1990, recomendaban la lidocaína tras el IAM, a pesar de que todos los ECA desde 1970 mostraban una tendencia a disminuir la supervivencia de los pacientes con IAM<sup>19</sup>. Antman et al utilizaron una forma de RS, denominado metaanálisis acumulativo, para conocer la eficacia de los diversos procedimientos terapéuticos tras el IAM y la fecha en que ésta estaba disponible.

La RS sigue un método reproducible por cualquier otro investigador, agrupa los datos de las investigaciones originales y aplica estrategias científicas con el fin de minimizar los sesgos y favorecer su reproducción. La RS sigue una serie de etapas cuya secuencia se expone en la tabla I<sup>20</sup>. El primer paso consiste en exponer los objetivos de la revisión, si la pregunta que se quiere responder no está bien formulada es probable que la revisión no sea pertinente. A continuación debe identificarse el mayor número de ECA que se encuentren publicados. Para ello la búsqueda

electrónica no es suficiente pues no garantiza la identificación de toda la información disponible. Un buscador inexperto en Medline (base de datos automatizada del Index Medicus, Biblioteca Nacional de Medicina, EE.UU.) sólo encontrará el 15 % de toda la información, un experto alcanza a recuperar el 45 % de todo lo disponible en la base de datos respecto a un determinado tema<sup>21-23</sup>. Posteriormente evaluaremos la calidad de cada estudio usando los mismos criterios de validez que para las publicaciones originales: asignación aleatoria, enmascaramiento, diseño del estudio, intención de tratar, etc., pues la revisión de peritos (*peer review*) de las revistas no garantiza la calidad de la investigación que se publica de ellas. Los criterios para incluir o excluir ECA mejora la homogeneidad de los resultados y disminuye el sesgo en la selección de artículos que apoyan conclusiones previas. Otro paso fundamental en la RS es la búsqueda exhaustiva con el fin de no perder ningún ECA relevante, para ello los contactos personales con los investigadores de los estudios originales ayudan a identificar los ECA realizados, estén o no publicados. La omisión de estos últimos aumenta la posibilidad de sesgo de publicación.

La RS es una técnica científica eficiente. Aunque muchas veces es laboriosa y ocupa mucho tiempo, este procedimiento es generalmente más rápido y menos costoso que embarcarse en un nuevo estudio. Un ejemplo de los beneficios de la RS nos lo enseñan Antman et al en un metaanálisis acumulativo de 33 ECA donde comparan el efecto de la estreptocinasa intravenosa con el placebo o con ningún tratamiento, en el IAM. Veinticinco de éstos fueron, en términos de mortalidad, favorables a la estreptocinasa, pero sólo en seis de ellos se había conseguido alcanzar una significación estadística. Los resultados, presentados como si se realizase un metaanálisis cada vez que se notificasen los resultados de un nuevo ECA, muestran que el efecto del tratamiento ya era estadísticamente significativo en el año 1973, a pesar de lo cual se siguieron realizando ECA, incluyendo los grandes multicéntricos GISSI-I, en 1986, e ISIS-2, en 1988<sup>24</sup>. Si se hubiese practicado esta revisión 20 años antes, se hubiese generalizado entonces el tratamiento trombolítico tras el IAM y se hubiesen salvado miles de vidas.

Las RS cuantitativas o metaanálisis aumentan el poder estadístico de los estudios originales, lo que es especialmente ventajoso en condiciones de baja prevalencia o cuando se evalúen efectos de pequeña magnitud<sup>17</sup>. Un ejemplo es el logotipo de la Colaboración Cochrane (fig. 1), que ilustra la capacidad de los corticoides en reducir

TABLA I  
ETAPAS EN LA ELABORACIÓN  
DE UNA REVISIÓN SISTEMÁTICA

1. Establecer los objetivos de la revisión y delimitar los criterios de selección
2. Búsqueda de los estudios que reúnan los criterios de selección
3. Tabular las características de cada estudio identificado y examinar su calidad metodológica
4. Aplicar los criterios de selección y justificar las exclusiones
5. Reunir el mayor número posible de conjunto de datos con incorporación de los investigadores de los ECA, si esto es posible
6. Analizar los resultados. Si es posible hacer una síntesis estadística de éstos (metaanálisis)
7. Realizar análisis de sensibilidad, si es apropiado y posible
8. Preparar un informe estructurado de la revisión, exponiendo los objetivos, describiendo los materiales y métodos, exponiendo los resultados y extrayendo las conclusiones

ECA: ensayo clínico aleatorizado.

la mortalidad perinatal en los recién nacidos de mujeres con riesgo de parto prematuro, mediante un metaanálisis que incluye 7 ECA. El efecto beneficioso es estadísticamente significativo, a pesar de que en sólo dos de los ECA mostraban un claro efecto a favor de los corticoides.

### Búsqueda de ECA relevantes

La identificación completa de ECA, incluso de aquellos publicados íntegramente, es realmente difícil. Reunir los ECA realizados en los últimos 40 a 50 años supone una labor ingente. Se calcula que se habrán realizado cientos de miles de ECA sobre diferentes aspectos de la asistencia sanitaria desde 1948, año del primer ECA publicado<sup>25</sup>. A esto se añade la falta de claridad en la descripción de los métodos y, el escollo que supone el idioma cuando el artículo viene publicado en otro que no sea el inglés o el del propio investigador. En este sentido Dickersin et al encontraron que la quinta parte de los ECA sobre visión identificados en Medline no son ingleses<sup>26</sup>.

La búsqueda exhaustiva en la bibliografía biomédica supone dos tipos de métodos: uno electrónico y otro manual. Por un lado, está la identificación de los ECA a través de las bases bi-

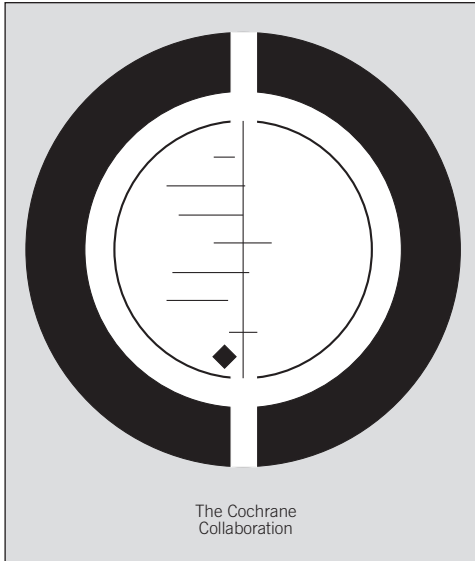


Fig. 1. Logotipo de la Colaboración Cochrane. Ilustra el resultado de agrupar los datos de 7 estudios diferentes (las líneas horizontales representan los intervalos de confianza correspondientes). El efecto global (rombo) es estadísticamente significativo, indicado porque está a la izquierda de la línea vertical (odds ratio = 1) y no contacta con ésta.

biográficas más comunes, de las cuales la más extendida es Medline, con 3.000 revistas indexadas, lo que supone el 15 % de todas las revistas biomédicas del mundo. Otras bases informatizadas pueden complementar la búsqueda en la bibliografía: Embase, de Excerpta Medica, incorpora un mayor número de revistas europeas, SCISEARCH, de Science Citation Index, incluye datos de citaciones, y bases de datos específicas como son Ageline, Psyclist. Al ser Medline la más universal de todas las bases informatizadas, es donde más se ha estudiado el rendimiento de las diferentes estrategias de búsqueda (tabla II), diseñadas para que sean más sensibles que específicas, pues es preferible acumular artículos falsamente identificados como ECA que perder un solo ECA<sup>27</sup>. Las principales limitaciones de la búsqueda electrónica son: su cobertura, Medline sólo incluye los artículos a partir de 1966, su rendimiento restringido por un índice imperfecto que hace que se recupere menos del 50 % de la información relevante a pesar de que ésta se encuentra en la base de datos<sup>28</sup>, y la exclusiva inclusión de revistas, eliminando una fuente importante de publicaciones, incluidas las que los anglosajones denominan “bi-

bliografía opaca” (*grey literature*) por ser de laborioso acceso a ella, como son las tesis doctorales o las comunicaciones a congresos. La otra forma de búsqueda en la bibliografía biomédica es la manual, consistente en el repaso pormenorizado de todos los números de las revistas elegidas en un período determinado, así como la revisión de la “bibliografía opaca”. Esto, aparte de ser una labor tediosa, es un trabajo ilimitado pero en donde pueden participar voluntarios no expertos en el área de la salud, obteniendo unos resultados muy satisfactorios: identificación del 90 % de ECA relevantes y un 20 % de falsos positivos<sup>29</sup>. Ha habido tentativas para localizar ECA no publicados mediante encuestas a un gran número de investigadores, pero los resultados han sido tan pobres que no ha compensado seguir haciéndolas<sup>30</sup>. La búsqueda manual es el método más seguro para la identificación de artículos, pero es un procedimiento que ocupa mucho tiempo. Se utiliza como referencia principal (*gold standard*) para comparar el rendimiento de cualquier otra técnica de búsqueda bibliográfica. Jadad y McQuay, comparando el rendimiento del método manual con Medline para la identificación de ECA sobre analgesia en 9 revistas de anestesia en 3 años (1970, 1980 y 1990), encontraron una precisión del 2,7 % y una sensibilidad del 99 frente al 52 y 87 %, respectivamente, con un tiempo para la identificación de un ECA de 4 min mediante el método manual y 24 seg mediante Medline. La Colaboración Cochrane (CC) fue una idea surgida para resolver el manejo y extracción relevante de toda la bibliografía biomédica.

### Colaboración Cochrane

Como respuesta a la llamada de Archie Cochrane de realizar, actualizar y difundir revisiones sistemáticas de ECA sobre los efectos de la asistencia sanitaria, se reunió en Oxford un grupo de investigadores con el objetivo de desarrollar un registro de ECA en el campo de la obstetricia<sup>31</sup>. Este fue el embrión de la Colaboración Cochrane (CC). Se estableció el primer centro Cochrane, el U.K. Cochrane Centre, hace 3 años en una antigua factoría de pan en un suburbio de esta ciudad universitaria. Durante este tiempo esta organización ha crecido hasta conformar una red internacional de instituciones e individuos comprometidos en la revisión sistemática y periódica de ECA<sup>32</sup> que se conoce como CC. En su corta existencia esta organización tiene 8 centros repartidos por todo el mundo (Reino Unido, Países Escandinavos, Canadá, EE.UU., Italia y Australia) y más de 1.000 investigadores colaborando en

TABLA II  
ESTRATEGIA PARA LA IDENTIFICACIÓN DE ENSAYOS CLÍNICOS ALEATORIZADOS  
POR MEDIO DE MEDLINE ACONSEJADA POR LA COLABORACIÓN COCHRANE

<i>Paso de búsqueda</i>	<i>Término buscado o combinación de pasos</i>
#1	RANDOMIZED-CONTROLLED-TRIAL in PT
#2	RANDOMIZED-CONTROLLED-TRIALS
#3	RANDOM-ALLOCATION
#4	DOUBLE-BLIND-METHOD
#5	SINGLE-BLIND-METHOD
#6	#1 or #2 or #3 or #4 or #5
#7	TG-ANIMAL not (TG = HUMAN and TG = ANIMAL)
#8	#6 not #7
#9	CLINICAL-TRIAL in PT
#10	explode CLINICAL-TRIALS
#11	(clin* near trial*) in TI
#12	(clin* near trial*) in AB
#13	(singl* or doubl* or trebl* or tripl*) near (blind* or mask*)
#14	(13 in TI) or (13 in AB)
#15	PLACEBOS
#16	placebo* in TI
#17	placebo* in AB
#18	random* in TI
#19	random* in AB
#20	RESEARCH-DESIGN
#21	#9 or #10 or #11 or #12 or #14 or #15 or #16 or #17 or #19 or #20
#22	TG = ANIMAL not (TG = HUMAN and TG = ANIMAL)
#23	#21 not #22
#24	#23 not #8
#25	TG = COMPARATIVE-STUDY
#26	explode EVALUATION-STUDIES
#27	FOLLOW-UP STUDIES
#28	PROSPECTIVE-STUDIES
#29	control* or prospectiv* or volunteer*
#30	(#29 in TI) or (#29 in AB)
#31	#25 or #26 or #27 or #28 or #30
#32	TG = ANIMAL not (TG = HUMAN and TG = ANIMAL)
#33	31 not 32
#34	33 not (8 or 24)

grupos de revisión internacionales, que abarcan 50 áreas clínicas (tabla III). La CC ha descubierto que más de la mitad de los 40.000 ECA publicados hasta 1985 no son recuperados en Medline ni por expertos documentalistas. De ahí que la colaboración proponga la búsqueda como manual como un complemento indispensable de la búsqueda electrónica en los índices automatizados de las principales revistas biomédicas. La Colaboración ha seguido el modelo del NPEU, de tal forma que las revisiones las realizan unos pocos grupos, cada uno de los cuales está formado por un reducido número de individuos que trabajan juntos en un área de interés común, coordinados por un equipo director. Dichos grupos pueden dedicarse a revisar ciertos tipos de enfermeda-

des<sup>33</sup>, variedades de tratamientos o especialidades completas<sup>34</sup>. La CC proporciona apoyo organizativo para reducir el trabajo a los revisores individuales y evitar la duplicación de esfuerzo.

La creación de un registro internacional de ECA iniciado ya por la CC ayudará a los revisores en la laboriosa tarea de la identificación de todos los ECA relevantes en una determinada revisión. Mientras, los coordinadores de los grupos de revisión ya han comenzado a elaborar guías y protocolos que aseguren la uniformidad y el buen hacer de las revisiones.

Para asegurar que los resultados del trabajo de la Colaboración sean ampliamente difundidos, cada grupo contribuye con su revisión a crear la Base de Datos Cochrane de RS (BDCRS), sobre

TABLA III  
ÁREAS SOMETIDAS A REVISIÓN POR LA COLABORACIÓN COCHRANE

Modificación del comportamiento Cáncer Salud infantil Medicina complementaria Dermatología Países en vías de desarrollo Sistema de distribución de fármacos Geriátría Medicina de urgencia Gastroenterología Ginecología Nefrología Enfermería Nutrición Fisioterapia Atención primaria Psiquiatría Salud pública Reumatología Intervenciones sociales Cirugía Vacunas Obstetricia Infertilidad Cuidados neonatales Ictus	Infecciones respiratorias agudas Asma y EPOC Dolor lumbar Enfermedad cardiovascular en atención primaria Demencia Salud bucal y dental Diabetes Epilepsia Control de fertilidad Hipoacusia Enfermedad cardíaca Hepatología Incontinencia Enfermedad inflamatoria intestinal Trastornos de aprendizaje Dolor Cuidados intensivos Enfermedades parasitarias Enfermedad vascular periférica Trastornos psicológicos en atención primaria Esquizofrenia Traumatología Trastornos visuales Cicatrización de heridas
--	---

EPOC: enfermedad pulmonar obstructiva crónica.

TABLA IV  
ESTRUCTURA DE UNA REVISIÓN  
DE LA BASE DE DATOS COCHRANE  
DE REVISIONES SISTEMÁTICAS

Cubierta: título, nombre de los autores de la revisión, equipo director responsable del grupo, fuente de financiación
Informe estructurado: introducción, objetivos, material y métodos, resultados y discusión
Cita completa de todos los estudios revisados, tanto de los incluidos como de los excluidos y las razones de su exclusión
Tabulación de las características de los ECA incluidos junto a una valoración sobre la calidad metodológica de cada uno de ellos
Tabulación de los resultados de la revisión, con una síntesis estadística (metaanálisis) cuando sea posible y apropiado
Dirección de contacto para más información sobre ECA relevantes no publicados o sobre ECA todavía no finalizados

ECA: ensayo clínico aleatorizado.

la cual nadie dispone de derechos de autor y es de libre acceso. Cada revisión de la BDCRS tiene un formato estándar (fig. 2).

La difusión por métodos electrónicos de la BDCRS es la que más ventajas tiene, debido a su permanente actualización y corrección a medida que se disponga de evidencia nueva. La base de datos completa se distribuye en línea y en CD-ROM. Las bases de datos más pequeñas creadas a partir de la base de datos principal se recopilan y se publican en disquetes (fig. 3).

Es importante poder hacer correcciones y mejorar las revisiones a la luz de las buenas críticas. La CC está intentando crear un sistema interactivo por medio del cual las sucesivas versiones de cada revisión reflejen no sólo la incorporación de nuevos datos sino cualquier tipo de críticas.

A medida que la CC crezca será posible que la información sobre los efectos de la asistencia sanitaria esté disponible para todo el mundo, tanto para los gestores de la sanidad como para los propios usuarios. Esta información, junto con las necesidades de la población, la disponibilidad de los recursos y las prioridades son obligatorias para ofrecer los cuidados de salud más apropiados.

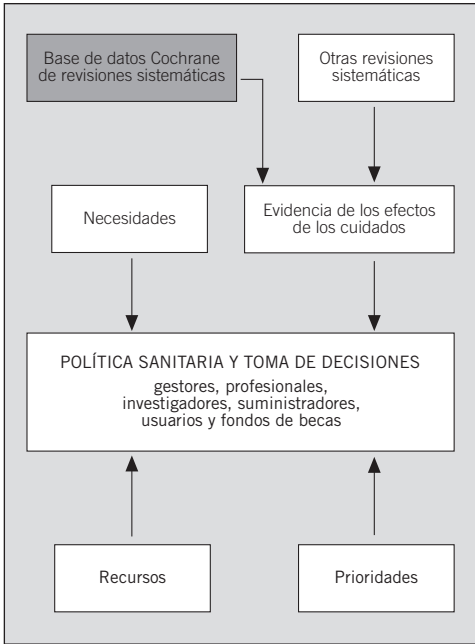


Fig. 2. Revisiones sistemáticas y actualizadas sobre los efectos sanitarios, esenciales, aunque no suficientes, para mejorar la política sanitaria y la toma de decisiones en la investigación y la atención sanitaria.

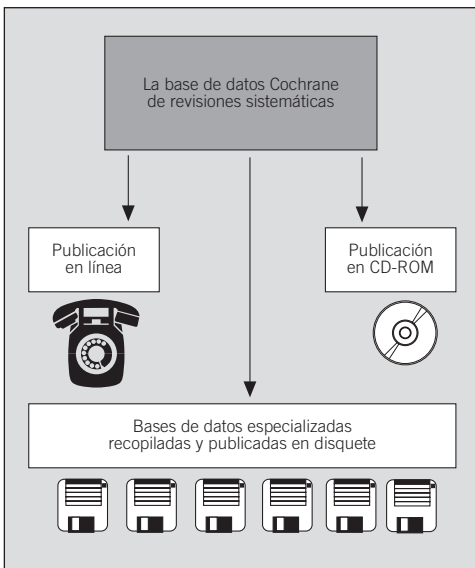


Fig. 3. Difusión por medios electrónicos de las revisiones sistemáticas.

## BIBLIOGRAFÍA

1. Chuang Stein C. The regression fallacy. *Drug Inf J* 1993; 27: 1.213-1.220.
2. Zayas Sanza JM. Regresión a la medida y evolución de la enfermedad. *Jano* 1995; 49: 1.120.
3. Meynert CL. *Clinical trials. Design, conduct and analysis*. Nueva York: Oxford University Press, 1986.
4. Cochrane AL. *Effectiveness and efficiency; random reflections on health services*. Londres: Nuffield Provincial Hospitals Trust, 1972.
5. Sackett D. The Cochrane Collaboration. En: Sackett D, editor. *The Cochrane Collaboration Handbook*. Section I. Oxford: Cochrane Collaboration, 1994.
6. Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. En: *Medicines for the year 2000*. Londres: Office of Health Economics, 1979; 1-11.
7. Haynes RB. How clinical journals could serve clinician readers better. En: Lock S, editor. *The Future of Medical Journals*. Londres: *Br Med J*, 1991; 116: 126.
8. Williamson JW, German PS, Weiss R, Skinner EA, Bowes F. Health science information management and continuing education of physicians. *Ann Intern Med* 1989; 110: 151-160.
9. Gotzsche PC. Reference bias in reports of drug trials. *Br Med J* 1987; 295: 654-656.
10. Early Breast Cancer Trialists' Collaborative Group. Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *Lancet* 1992; 339: 1-15.
11. Advanced Ovarian Cancer Trialists' Group. Chemotherapy in advanced ovarian cancer: an overview of randomized clinical trials. *Br Med J* 1991; 303: 884-893.
12. Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged anti-platelet treatment. *Br Med J* 1988; 296: 320-331.
13. Selective Decontamination of the Digestive Tract Trialists' Collaborative Group. Meta-analysis of randomized controlled trials of selective decontamination of the digestive tract. *Br Med J* 1993; 307: 525-532.
14. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272: 1.926-1.931.
15. Cochrane AI. Foreword. En: Chalmers I, Enkin M, Keine MJNC, editores. *Effective care in pregnancy and childbirth*. Oxford: Oxford University Press, 1989.
16. Ad Hoc Group for critical Appraisal of the Medical Literature. Academic and clinic: a proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; 106: 598-604.
17. Mulrow CD. Systematic Reviews: Rationale for systematic reviews. *Br Med J* 1994; 309: 597-599.
18. Dickersin K, Berlin JA. Meta-analysis. State of the science. *Epidemiol Rev* 1992; 309: 597-599.
19. Antman EM, Lau J, Kupelnik B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *JAMA* 1992; 268: 240-248.

20. Improving policies and decisions in health care and research. The role of randomized controlled trials (RCTs). The Cochrane Centre. NHS R&D Programme 1993.
21. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Comparison of MEDLINE searching with a perinatal trials database. *Control Clin Trials* 1985; 6: 306-317.
22. Poynard T, Conn Ho. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. *Control Clin Trials* 1985; 6: 271-279.
23. Kirplani H, Schimdt B, McKibbin KA, Haynes RB, Sinclair JC. Searching MEDLINE for randomized clinical trials involving care of the newborn. *Pediatrics* 1989; 83: 543-546.
24. Antman EM, Lau J, Kupelnik B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *JAMA* 1992; 268: 240-248.
25. Bradford-Hill H. Streptomycin treatment of pulmonary tuberculosis: A Medical Research Council Investigation. *Br Med J* 1948; 2: 769-782.
26. Dickersin K, Scherer R, Lefebvre C. Identification of relevant studies for systematic reviews. *Br Med J* 1994; 309: 1.286-1.291.
27. Dickersin K. Establishing and maintaining an international register of RCTs. En: Sackett D, editor. *The Cochrane Collaboration Handbook*. Section V. Oxford: Cochrane Collaboration, 1994.
28. Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* 1993; 2: 39 paragraphs.
29. Jadad AR, Jamal S, Carroll D, McQuay HJ. Can volunteers with no research experience hand search medical journal effectively to find randomized controlled trials? *Control Clin Trials* 1994; 15: 62-73
30. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989; 84: 374-380.
31. Chalmers I, Enkin M, Keirse MJ, editores. *Effective care in pregnancy and childbirth*. Oxford: Oxford University Press, 1989.
32. Chalmers I, Haynes B. Reporting, updating, and correcting systematic reviews of the effects of health care. *Br Med J* 1994; 309: 862-865.
33. Early Breast Cancer Trialist's Collaborative Group. Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *Lancet* 1992; 339: 1-15.
34. Chalmers I, Enkin M, Keirse MJNC, editores. *Effective care in pregnancy*. Oxford: Oxford University Press, 1989.



---

# Identificación y descripción de los ensayos clínicos publicados en revistas españolas de medicina general e interna durante el período 1971-1995

---

Joan Martí<sup>a</sup>, Xavier Bonfill<sup>a</sup>, Gerard Urrutia<sup>a</sup>, Juan Ramón Lacalle<sup>b</sup> y Rafael Bravo<sup>c</sup>

<sup>a</sup>Centro Cochrane Español. Fundació Parc Taulí. Sabadell. Barcelona. <sup>b</sup>Departamento de Ciencias Socio-Sanitarias. Universidad de Sevilla. <sup>c</sup>Centro de Salud El Greco. Getafe. Madrid.

Una búsqueda global y no sesgada de la bibliografía biomédica es una de las diferencias claves entre una revisión sistemática y una revisión tradicional<sup>1</sup>. Disponer de las mejores evidencias con las que tomar decisiones terapéuticas, es decir, buenos ensayos clínicos controlados (ECC), es excesivamente difícil. En primer lugar, los autores de una revisión convencional no describen sus métodos de búsqueda con suficiente claridad como para permitir una valoración precisa de este proceso. En segundo lugar, el término ECC no fue incorporado a las bases de datos bibliográficas hasta 1990 y, por lo tanto, los ECC anteriores a esta fecha tuvieron que ser clasificados bajo categorías muy amplias, mezclados con otro tipo de estudios<sup>2</sup>. Finalmente, existen datos que sugieren que los descriptores disponibles en aquellas bases de datos no han sido utilizados siempre de manera consistente por las personas encargadas de su codificación<sup>1,2</sup>. Por lo tanto, la búsqueda realizada exclusivamente por medios electrónicos es insuficiente y se ha demostrado que ni siquiera el mejor documentalista con una estrategia de búsqueda óptima puede encontrar más de la mitad de los ensayos que contiene Medline<sup>3</sup>. No obstante, el problema de la indización confusa de los trabajos en las bases de datos electrónicas no es el único: éstas son de gran ayuda para superar muchas dificultades de acceso a la bibliografía, pero pueden haber creado la falsa impresión de que contienen toda la evidencia existente cuando en realidad no es así<sup>4</sup>. Las dos principales bases de datos electrónicas en ciencias de la salud, Medline y Embase, sólo abarcan

unas 6.000 de las aproximadamente 30.000 revistas que se publican en todo el mundo. Además, incluyen preferentemente revistas en inglés mientras que los trabajos publicados en revistas no anglófonas están mal representados, lo que es una fuente potencial de sesgos de publicación<sup>5-8</sup>.

Para asegurar la máxima fiabilidad posible, es necesario usar múltiples sistemas de identificación, que incluyan una estrategia de búsqueda manual rigurosa y exhaustiva, con el fin de identificar el máximo número posible de ECC<sup>1,9</sup>. Por búsqueda manual se entiende el examen de cada uno de los números de una determinada revista con lectura progresiva de todas sus secciones hasta que sea posible clasificar cada estudio como ECC o descartarlo como tal. De esta forma también puede revisarse la que en inglés se denomina "bibliografía gris" (*grey literature*) por ser de laborioso acceso a ella, como son las tesis doctorales y las comunicaciones a congresos. Esta tarea representa un trabajo imprescindible, sólo posible gracias a la colaboración de multitud de voluntarios, aunque no sean expertos en el área de la salud, y consigue unos resultados muy satisfactorios: identificación del 90 % de ECC relevantes y sólo un 20 % de falsos positivos<sup>10</sup>. Por tanto, en estos momentos la búsqueda manual representa un complemento indispensable a la búsqueda electrónica para identificar exhaustivamente todos los ensayos clínicos realizados.

Para hacer esta tarea de la manera más eficiente posible, la Colaboración Cochrane ha organizado una estrategia de búsqueda manual mundial, que incluye la creación de un registro internacional de ensayos clínicos con los resultados obtenidos en las búsquedas manuales de miles de revistas de todo el mundo, además de las búsquedas electrónicas en las principales bases de datos bibliográficas. Aquellos ECC pueden ser consultados en el Cochrane Controlled Trials Register, una de las bases de datos de la *Cochrane Library*<sup>11</sup>. En consonancia con este gran esfuerzo

---

Publicado en *Med Clin (Barc)* 1999; 112 (Supl 1): 28-34. Este estudio ha sido posible en parte gracias al proyecto FIS (Nº: 96/1802E) y a una ayuda del II Pla de Recerca de Catalunya 1997-2000, en el marco de la *Xarxa temàtica: "Medicina basada en l'evidència: la Col·laboració Cochrane"*. Además, Joan Martí es becario del Instituto de Salud Carlos III (1997).

internacional de cooperación, el Centro Cochrane Español (CCE) coordina la búsqueda de los ECC realizados en España<sup>12</sup>. Aparte de la revisión manual de todas las revistas en ciencias de la salud españolas, se pretende consultar Medline, Embase, la Base de Datos Española de Ensayos Clínicos del Registro de la Dirección General de Farmacia del Ministerio de Sanidad, los ensayos aprobados por los diferentes Comités Éticos de Investigación Clínica (CEIC) españoles y los promovidos desde laboratorios farmacéuticos e industrias similares establecidas en España.

Como primera parte de este ambicioso proyecto, se presentan en este artículo los resultados de la búsqueda manual de los ECC publicados en revistas españolas de medicina general e interna, desde 1971 hasta 1995 (25 años). También se compara el rendimiento de la búsqueda manual y de la electrónica cuando se utiliza Medline.

## Métodos

### Búsqueda manual

*Identificación de las publicaciones.* Para la identificación de los ECC publicados en las revistas españolas de medicina general e interna fue necesario conocer en primer lugar el total de revistas que debían ser sometidas al proceso de revisión manual. Se han considerado susceptibles de revisión las que han publicado estudios originales en el ámbito de la medicina general e interna (elegidas según el mejor criterio de los autores, tabla I) durante el período 1971-1995.

1. *Índice Médico Español (IME).* El Índice Médico Español (IME), iniciado en 1964 en Valencia, y actualmente bajo la dependencia del Instituto de Estudios Documentales e Históricos de Valencia y del Consejo Superior de Investigaciones Científicas, es la principal base de datos bibliográfica específicamente española en ciencias de la salud. Además, desde 1971 sus datos están informatizados en Knosys, la base de datos electrónica del IME, y su consulta es mucho más fácil, motivo por el cual el inicio del período de revisión para este proyecto se sitúa en aquel año. El IME permite, por lo tanto, la posibilidad de identificar las principales publicaciones en medicina editadas en el país, aunque no todas.

2. *Catálogo Nacional de Publicaciones Periódicas en Bibliotecas de Ciencias de la Salud Españolas C-17.* Editado por el Centro de Información y Documentación Científica (CINDOC) en CD-ROM en el año 1996. Permite consultar elec-

trónicamente los títulos de las revistas que contiene cada una de las 420 bibliotecas de ciencias de salud española.

3. *Catálogos colectivos de publicaciones periódicas en ciencias de la salud.* Editados en papel por algunas Comunidades Autónomas, representan la posibilidad de conocer los títulos de las revistas de sus respectivas bibliotecas. Se han revisado los catálogos de Cataluña, Comunidad Valenciana, Baleares, Andalucía y Madrid.

4. *Otros: notificaciones personales espontáneas,* que han servido para completar la lista de revistas.

*Criterios de inclusión de los ensayos clínicos controlados.* Para ser considerado un ECC, un estudio debía cumplir los siguientes criterios: a) comparar tratamientos en seres humanos; b) ser prospectivo (las intervenciones debían haber sido planificadas antes de que tuviera lugar el estudio y la asignación de los sujetos a cada intervención tenía que ser realizada por los investigadores; c) compararse entre sí dos o más tratamientos o intervenciones (uno podía ser un grupo control sin tratamiento o un placebo). Las intervenciones podían ser de cualquier tipo: farmacológicas, quirúrgicas, diagnósticas, educativas, rehabilitadoras, organizativas u otras, y d) tener un método aleatorio de asignación a los tratamientos. Las unidades de aleatorización podían ser individuos, grupos (hospitales, comunidades), órganos (p. ej., ojos) o partes del cuerpo humano (p. ej., dientes)<sup>13</sup>. Los estudios que cumplían estos criterios eran clasificados posteriormente en:

1. *Ensayos controlados con distribución aleatoria (ECDA).* Cuando el estudio cumplía los cuatro criterios de elegibilidad y el/los autor/es explicitaban (habitualmente usando alguna variante del término "aleatorio", "azar", "randomizado", para describir el método de asignación usado) que los grupos comparados se crearon mediante asignación aleatoria.

2. *Ensayos clínicos controlados (ECC).* Cuando un ensayo elegible no había sido descrito explícitamente como "aleatorio" o similar. Este término es utilizado para distinguir los posibles ECDA y los casi ECDA de los "verdaderos" ECDA. Procedimientos "casi al azar" son la utilización de monedas (cara o cruz), los números de historia clínica, los días de la semana, la secuencia de visitas y otros.

*Identificación de los ensayos clínicos.* El proceso de búsqueda manual ha sido sistemático

TABLA I  
REVISTAS ESPAÑOLAS DE MEDICINA GENERAL E INTERNA IDENTIFICADAS

Número	Nombre de la revista	Periodo revisado	Posibles ECC	ECC	Revisor
1	Acta Médica de Tenerife	1971-1978	1	0	JM
2	Acta Médica del Distrito Universitario de Salamanca	1972-1975	2	1	AM, JM
3	Actas Clínicas Delfos	1984-1985	0	0	JM
4	Actualidad Médica	1971-1995	1	0	AM, JR, JM
5	Anales de la Facultad de Medicina. Universidad de Salamanca	1971-1977	0	0	JM
6	Anales de la Real Academia de Medicina y Cirugía de Salamanca	1971-1994	0	0	AM
7	Anales de la Real Academia de Medicina y Cirugía de Cádiz	1971-1995	0	0	AM
8	Anales de la Real Academia Nacional de Medicina	1971-1995	1	0	AM, JM
9	Anales de la Universidad Hispalense	1971	0	0	JR
10	Anales de la Universidad de Murcia, Medicina	1977-1978	0	0	AM
11	Anales de Medicina (Barcelona, 1973)	1973-1975	1	1	JM
12	Anales de Medicina de Sevilla	1971-1995	0	0	JR
13	Anales de Medicina Interna	1983-1995	40	25	AM, JM
14	Anales de Medicina y Cirugía	1971-1982	1	0	JM
15	Anales de Medicina, Medicina	1971-1972	2	0	JM
16	Anales del Hospital de la Sta. Cruz y S. Pablo	1971-1975	1	0	JM
17	Annals de Medicina	1983-1995	96	94	JM, SL, CE
18	Archivos de Estudios Médicos Aragoneses	1971-1984	0	0	EB
19	Archivos de la Facultad de Medicina de Madrid	1971-1981	3	1	JM, RB, CB
20	Archivos de la Facultad de Medicina de Oviedo	1976-1995	0	0	AM, JM, RB, CB
21	Archivos de la Facultad de Medicina de Zaragoza	1971-1995	15	9	JM, AM
22	Archivos del Hospital Provincial de San Juan de Dios de Málaga	1980-1983	0	0	VA
23	Atención Médica	1989-1995	0	0	JM
24	Atención Primaria	1984-1995	41	30	IF
25	Boletín de la Fundación Jiménez Díaz	1971-1984	7	3	AM, RB, CB, JM
26	Boletín de Patología Médica	1971-1975	0	0	AM, JM
27	British Medical Journal (ed. esp.)	1986-1995	0	0	AM
28	Cuadernos de Atención Primaria	1993-1995	2	1	MV
29	Canarias Médica	1985-1992	0	0	JM, JR
30	Centro de Salud	1993-1995	3	1	JM
31	Centro Médico	1979-1985	0	0	JM
32	Ciencia Médica	1984-1995	9	5	AM, JM
33	Clinic (Barcelona)	1980-1989	0	0	JM
34	Clinica Rural	1971-1995	2	0	AM, JM
35	Cuadernos de Salud	1989-1991	0	0	JM

(Continua)

TABLA I  
REVISTAS ESPAÑOLAS DE MEDICINA GENERAL E INTERNA IDENTIFICADAS (Continuación)

Número	Nombre de la revista	Periodo revisado	Posibles ECC	ECC	Revisor
36	Die Medizinische Welt (ed. esp.)	1975-1979	4	4	JM, AM
37	Europa Médica	1972-1975	0	0	JM
38	Gaceta Médica de Bilbao	1971-1995	6	3	JM, AM
39	Gaceta Clínica	1971-1995	8	1	JM, AM, ES
40	Hispalis Médica	1980-1992	0	0	JR
41	JAMA (ed. esp.)	1975-1981/1992-1995	0	0	AM
42	L'Internista	1978-1983	0	0	JM
43	Lancet (ed. esp.)	1982-1995	14	14	AM
44	Mapfre Medicina	1990-1995	0	0	MV
45	Medicina Asturiana	1971-1973	0	0	JM, MV
46	Medicina Balear	1986-1994	0	0	AM, EG
47	Medicina Clínica	1971-1995	196	93	JM, AM
48	Medicina de Madrid	1974-1976	8	7	RB, CB
49	Medicina Española	1971-1985	25	7	JM
50	Medicina Galaica	1978-1990	0	0	MV, AM
51	Medicina Integral	1980-1995	0	0	AM, JM
52	Medicina Militar	1983-1995	11	8	JM, FG
53	Medicine (Madrid)	1976-1986	0	0	AM
54	MEDIFAM	1991-1995	2	1	JV
55	Medizinische Klinik (ed. esp.)	1971-1983	39	24	JM
56	MTA Medicina Interna	1983-1995	0	0	AM
57	New England Journal of Medicine (ed. esp.)	1971-1977	3	1	AM
58	Nuevos Archivos de la Facultad de Medicina	1982-1985	4	2	RB, AM, CB, JM
59	Práctica Internista	1971	0	0	JM
60	Revista Clínica Española	1971-1995	174	67	AM, JM
61	Revista de la Reial Acadèmia de Catalunya	1992-1995	0	0	CE, SL
62	Revista de la Facultad de Medicina de Sevilla	1971-1974	4	0	JR, JM
63	Revista de la Reial Acadèmia de Barcelona	1986-1991	0	0	CE, JM
64	Revista de Medicina de la Universidad de Navarra	1971-1995	16	10	JM, AM
65	Revista Iberoamericana de Investigación Clínica	1982-1984	10	4	JM
66	Salud Rural	1983-1995	2	0	AM, JM
67	Sant Pau	1980-1989	0	0	JM
68	SEMER	1977-1995	13	7	AM, JM

para cada uno de los números de las publicaciones identificadas. En esta tarea han participado 15 revisores (que se especificarán posteriormente en el anexo 1; tabla I). Los revisores manuales han sido personas que han ofrecido desinteresadamente su colaboración al Centro Cochrane español (CCE), además de poder contar con una persona contratada durante algunos períodos. Para formar a los futuros revisores se realizó una prueba piloto, consistente en la revisión manual del período de una revista ya revisada desde el CCE. Para llevar a cabo esta tarea homogéneamente se editó una *Guía de búsqueda manual* (disponible en nuestra página web: <http://www.altaveu.chpt.es/cochrane/>), adaptación de la utilizada por la Colaboración Cochrane en todo el mundo. En ésta se establece que cada revista debe ser revisada detenidamente, de principio a fin, no sólo los artículos originales sino también las cartas al director y las referencias bibliográficas. Las etapas recomendadas son: a) lectura del índice de materias; b) localización de palabras clave en el título (p. ej., aleatorizado, ciego, azar, etc.); c) lectura del resumen, y d) lectura del apartado de material y métodos. Cada revisor manual se comprometía a realizar la revisión de una/s revista/s concretas (de un número determinado de volúmenes o de su totalidad), para cuya actividad disponía del asesoramiento de uno de los autores (JM).

#### *Búsqueda en Medline de ensayos clínicos controlados publicados en revistas españolas de medicina general e interna*

Para comprobar la recuperabilidad mediante Medline de los ECC publicados en revistas españolas de medicina general e interna se realizó una comparación de la búsqueda manual con la búsqueda electrónica (validada también por la Colaboración Cochrane y disponible en nuestra página web) en 4 revistas indizadas en dicha base de datos: *MEDICINA CLÍNICA*, *Revista Clínica Española*, *Atención Primaria* y *Anales de Medicina Interna*. Previamente, para desarrollar dicha tarea, se confirmó en la base de datos Serline la presencia de revistas españolas de medicina general e interna incluídas en Medline. Sólo se han indizado 45 revistas españolas en esta base de datos durante nuestro período de estudio; de ellas únicamente seis fueron de medicina general: *MEDICINA CLÍNICA*, *Revista Clínica Española*, *Atención Primaria*, *Anales de Medicina Interna*, *Revista de Medicina de la Universidad de Navarra* y *Anales de la Real Academia Nacional de Medicina* (tabla II).

#### *Extracción de la información*

Para realizar la recogida de información de una manera coordinada y completa se editaron dos tipos de cuestionarios. El primero de ellos sirvió para registrar las citas bibliográficas de los artículos identificados como posibles ECC, utilizándose un cuestionario para cada cita. El segundo estaba destinado a realizar el seguimiento de la revisión de la revista, para saber en todo momento los números que habían sido revisados y los que quedaban pendientes. Los diferentes revisores, a medida que finalizaban su tarea, enviaban al CCE los cuestionarios mencionados, desde donde se acudía a las diversas bibliotecas para fotocopiar los artículos. Una vez reproducidos, eran revisados (JM) acerca del cumplimiento de los criterios de inclusión y, una vez aceptados, eran clasificados entre RCT y CCT. Este proceso de recogida de la información permitía así a los revisores dedicarse íntegramente a las tareas de revisión manual. En la tabla III se detallan las variables recogidas de cada ECC.

#### *Análisis*

Se creó una base de datos con el programa Microsoft-Access®, a partir de la cual se calcularon las frecuencias de las diferentes variables de interés de cada uno de los ECC identificados. Para estudiar la evolución en el número de ECC publicados se establecieron intervalos de 5 años. Para analizar las frecuencias del número de participantes en los ECC se crearon 3 categorías: ≤ 50 participantes, entre 51 y 200 participantes y más de 200 participantes. Para observar la distribución en la realización de los ECC en toda España se agruparon por Comunidades Autónomas. La base de datos mencionada puede consultarse en nuestra página web y va a permitir múltiples explotaciones en el futuro.

## **Resultados**

#### *Identificación de las publicaciones*

Se identificaron un total de 68 revistas de medicina general e interna, mediante las siguientes fuentes de información utilizadas secuencialmente: a) IME: 29 revistas de medicina general indizadas y 12 más no indizadas en esta base de datos (sólo se identificaba el título de la publicación), en total 41 (60,3%); b) C-17: 9 revistas (13,2%); c) catálogos colectivos de publicaciones periódicas en ciencias de la salud: 14 revis-

TABLA II  
REVISTAS ESPAÑOLAS INCLUIDAS EN MEDLINE DURANTE EL PERÍODO DE ESTUDIO (1971-1995)

Número	Nombre de la revista	Abreviación de la indexación en Medline	Período de indexación
1	Acta-Obstétrica-y-Ginecológica-Hispano-Lusitana	Acta-Obstet-Ginecol-Hisp-Lusit	1971-1980
2	Acta-Otorrinolaringológica-Española	Acta-Otorrinolaringol-Esp	1988-1995
3	Actas-Dermosifilográficas	Actas-Dermosifilogr	1971-1981
4	Actas-Luso-Españolas-de-Neurología-y-Psiquiatría	Actas-Luso-Esp-Neurol-Psiquiatr	1971
5	Actas-Luso-Españolas-de-Neurología-Psiquiatría-y-Ciencias-Afines	Actas-Luso-Esp-Neurol-Psiquiatr-Cienc-Afines	1973-1995
6	Actas-Urológicas-Españolas	Actas-Urol-Esp	1973-1995
7	Allegología-et-Immunopatología	Allergol-Immunopathol-Madr	1971-1976
8	Anales-Españoles-de-Odontomatología	An-Esp-Odontomatol	1974-1995
9	Anales-Españoles-de-Pediatría	An-Esp-Pediatr	1988-1995
10	Anales-de-Medicina-Interna	An-Med-Interna	1988-1995
11	Anales-Otorrinolaringológicos-Iberoamericanos	An-Otorrinolaringol-Ibero-Am	1974-1995
12	Anales-de-la-Real-Academia-Nacional-de-Medicina	An-R-Acad-Nac-Med-Madr	1971-1995
13	Archivos-de-Bronconeumología	Arch-Bronconeumol	1994-1995
14	Archivos-de-Odontomatología	Arch-Odontol-Estomatol	1985-1988
15	Archivos-Españoles-de-Urología	Arch-Esp-Urol	1977-1995
16	Atención-Primaria	Aten-Primaria	1989-1995
17	Avances-en-Odontomatología	Av-Odontomatol	1985-1991
18	Avances-Periodoncia	Av-Periodoncia	1989
19	Cirugía-Pediátrica	Cir-Pediatr	1988-1995
20	Divulgación-Cultural-Odontológica	Divulg-Cult-Odontol	1971-1972
21	Enfermedades-Infecciosas-y-Microbiología-Clinica	Enferm-Infec-Microbiol-Clin	1989-1995
22	Gaceta-Sanitaria	Gac-Sanit	1987-1995
23	Gastroenterología-y-hepatología	Gastroenterol-Hepatol	1995
24	Histology-and-histopathology	Histol-Histopathol	1986-1995
25	International-Journal-of-Developmental-Biology	Int-J-Dev-Biol	1989-1995
26	Journal-of-Investigational-Allergy-and-Clinical-Immunology	J-Investig-Allergol-Clin-Immunol	1990-1995
27	Medicina-Clinica	Med-Clin-Barc	1978-1995
28	Methods-and-Findings-in-Experimental-and-Clinical-Pharmacology	Methods-Find-Exp-Clin-Pharmacol	1979-1995
29	Microbiología	Microbiologia	1985-1995
30	Neurología	Neurologia	1986-1995
31	Nutrición-Hospitalaria	Nutr-Hosp	1989-1995
32	Revista-Clinica-Española	Rev-Clin-Esp	1971-1995
33	Revista-Española-de-Anestesiología-y-Reanimación	Rev-Esp-Anestesiol-Reanim	1971-1995
34	Revista-Española-de-Cardiología	Rev-Esp-Cardiol	1971-1995
35	Revista-Española-de-Enfermedades-Digestivas	Rev-Esp-Enferm-Dig	1990-1995
36	Revista-Española-de-Fisiología	Rev-Esp-Fisiol	1971-1995
37	Revista-Española-de-Obstetricia-y-Ginecología	Rev-Esp-Obstet-Ginecol	1971-1979
38	Revista-Española-de-Salud-Pública	Rev-Esp-Salud-Publica	1995
39	Revista-Ibero-Americana-de-Ortodoncia	Rev-Iberoam-Ortod	1981-1989
40	Revista-de-Enfermería	Rev-Enferm	1980-1995
41	Revista-de-Medicina-de-la-Universidad-de-Navarra	Rev-Med-Univ-Navarra	1971-1995
42	Revista-de-Neurología	Rev-Neurol	1995
43	Revista-de-Sanidad-e-Higiene-Pública	Rev-Sanid-Hig-Publica-Madr	1971-1995
44	Sangre	Sangre-Barc	1971-1995
45	Toko-Ginecología-Práctica	Tokoginecol-Pract	1971-1977

TABLA III  
VARIABLES VALORADAS EN LOS  
DIFERENTES ENSAYOS CLÍNICOS  
CONTROLADOS

1. Tipo de publicación
Artículo
Comunicación a congreso
Carta
Otros (comunicación previa, etc.)
2. Año de realización y publicación
3. Problema de salud investigado
4. Intervenciones estudiadas
5. Fase del estudio (I, II, III o IV)
6. Número de participantes
7. Método de asignación aleatoria
ECC
ECDA
8. Tipo de enmascaramiento
Doble ciego
Simple ciego
9. Tiempo de seguimiento (en meses)
10. Resultados (outcomes)
11. Nacional (por Comunidad Autónoma) o internacional
12. Multicéntrico
13. Ámbito de realización
Atención primaria
Atención hospitalaria
Ambas
Otros
14. Tipo de promoción
Industria farmacéutica
Agencias de Financiación de Investigación
Otros

ECDA: ensayos controlados con distribución aleatoria; ECC: ensayos clínicos controlados.

tas (20,6%), y d) otras vías informales de consulta: 4 revistas (5,9%) (tabla I).

#### Búsqueda de ensayos clínicos controlados españoles en Medline

La estrategia de búsqueda exclusiva electrónica con las 4 revistas citadas españolas de medicina general indizadas en Medline demostró la pérdida de un 32% de ensayos clínicos (tabla IV). Mediante búsqueda manual se identificaron un total de 215 ECC correspondientes al período 1971-1995, mientras que a través de la búsqueda electrónica se han podido identificar sólo 146. La sensibilidad de la búsqueda en Medline (proporción del número total de ensayos clínicos identificados en la búsqueda manual que son recuperados en Medline) fue del 68% y la precisión

TABLA IV  
COMPARACIÓN DE LA BÚSQUEDA  
MANUAL CON LA BÚSQUEDA EN  
MEDLINE DE ENSAYOS CLÍNICOS  
CONTROLADOS (ECC) EN 4 REVISTAS  
ESPAÑOLAS DE MEDICINA GENERAL  
INDIZADAS

	Medline		
	Sí	No	Total
Búsqueda manual			
Sí	146	69	215
No	0	0	0
Total	146	69	215

Sensibilidad = 68% (146 ECC identificados a través de Medline/215 total de ECC)

Precisión = 4,1% (146/3.570 posibles ECC identificados)

ECC no recuperables vía Medline = 32,1%

MEDICINA CLÍNICA: 93 ECC; Revista Clínica Española: 67 ECC; Atención Primaria: 30 ECC; Anales de Medicina Interna: 25 ECC.

(proporción de artículos encontrados en Medline que corresponden a ensayos clínicos) fue del 4,1% (se recuperaron como posibles ECC 1.465 documentos publicados en MEDICINA CLÍNICA, 1.045 en *Revista Clínica Española*, 805 en *Atención Primaria* y 255 en *Anales de Medicina Interna*).

#### Características de los ensayos clínicos controlados

Se identificaron un total de 767 posibles ECC, entre 1971 y 1995, que una vez validados se convirtieron en 417 ECC publicados en las revistas mencionadas (tabla V). El primer ECC de este período apareció a finales de 1971 en la revista MEDICINA CLÍNICA (Gibert J, Rodés J. Acción terapéutica de la asociación de nucleopéptidos hepáticos y coenzima A en la hepatitis. *Med Clin (Barc)* 1971; 57: 1.007-1.012) y el primer ECDA apareció en esta misma revista en 1972 (Padrós W. Estudio doble ciego de la acción clínica del dobesilato de calcio sobre los síndromes de insuficiencia venosa. *Med Clin (Barc)* 1972; 58: 515-519). A partir de entonces se fue incrementando progresivamente el número de ensayos publicados (fig. 1), aunque en la mayoría de ellos, 364 (87%), no se informaba de la fecha de realización. El número de revistas que publicaron algún ECC durante el período de estudio fue de 26 (38,2%), seis de ellas indizadas en Medline y 14 en Knosys

TABLA V  
REVISTAS QUE HAN PUBLICADO ALGÚN ENSAYO CLÍNICO CONTROLADO  
DURANTE EL PERÍODO 1971-1995

Número	Revista	Número de ECC publicados	ECC	ECDA
1	<i>Annals de Medicina</i>	94	92	2
2	<i>Medicina Clínica</i>	93	78	15
3	<i>Revista Clínica Española</i>	67	65	2
4	<i>Atención Primaria</i>	30	28	2
5	<i>Anales de Medicina Interna</i>	25	21	4
6	<i>Medizinische Klinik (ed. esp.)</i>	24	21	3
7	<i>Lancet (ed. esp.)</i>	14	11	3
8	<i>Revista de Medicina de la Universidad de Navarra</i>	10	8	2
9	<i>Archivos de la Facultad de Medicina de Zaragoza</i>	9	9	0
10	<i>Medicina Militar</i>	8	8	0
11	<i>Medicina Española</i>	7	7	0
12	SEMER	7	6	1
13	<i>Ciencia Médica</i>	5	5	0
14	<i>Die Medizinische Welt (ed. esp.)</i>	4	4	0
15	<i>Revista Iberoamericana de Investigación Clínica</i>	4	4	0
16	<i>Boletín de la Fundación Jiménez Díaz</i>	3	2	1
17	<i>Gaceta Médica de Bilbao</i>	3	3	0
18	<i>Nuevos Archivos de la Facultad de Medicina</i>	2	2	0
19	<i>Acta Médica de Distrito Universitario de Salamanca</i>	1	1	0
20	<i>Anales de Medicina (Barcelona, 1973)</i>	1	1	0
21	<i>Archivos de la Facultad de Medicina de Madrid</i>	1	1	0
22	<i>Cuadernos de Atención Primaria</i>	1	0	1
23	<i>Centro de Salud</i>	1	0	1
24	<i>Galicia Clínica</i>	1	1	0
25	MEDIFAM	1	1	0
26	<i>New England Journal of Medicine (ed. esp.)</i>	1	1	0
Total		417	380	37

ECC: ensayos clínicos controlados; ECDA: ensayos controlados con distribución aleatoria.

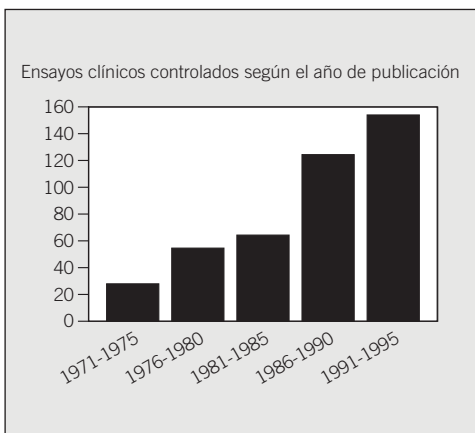


Fig. 1. Ensayos clínicos controlados según el año de publicación.

(la base de datos electrónica del IME). El número de revistas que publicaron ECC y no estaban indexadas en ninguna base de datos electrónica fue de siete: (*Acta Médica del Distrito Universitario de Salamanca*, *Anales de Medicina* (Barcelona, 1973), *Cuadernos de Atención Primaria*, *Die Medizinische Welt* (ed. esp.), *Medizinische Klinik* (ed. esp.), *The Lancet* (ed. esp.) y *New England Journal of Medicine* (ed. esp.).

Del total de ECC publicados 286 (68,6 %) lo fueron en forma de artículo original, mientras que 113 (27 %) fueron identificados como presentaciones a congresos, 16 (3,8 %) como cartas al director y el resto, dos (0,47 %), como comunicaciones previas. La mayoría de los ECC fueron realizados en España, 405 (97,1 %), mientras que 12 (3 %) fueron realizados en otros países, como Italia o Argentina. Por Comunidades Autónomas destacaron Cataluña con 163 (39 %), Madrid con



71 (17 %) y Andalucía con 31 (7,4 %), seguidas de la Comunidad Valenciana y de Euskadi con 17 (4,1 %) y 16 (3,8 %) ECC, respectivamente, y cabe destacar que en 22 (5,3 %) de los casos los ECC se han realizado en varias Comunidades Autónomas de forma simultánea. La información sobre el lugar de realización del ECC no puede identificarse en 36 (8,6 %) de los ECC publicados.

La mayoría de los ECC, 333 (79,7 %), se han desarrollado en hospitales mientras que sólo 42 (10 %) en centros de atención primaria y 21 (5 %) en universidades o laboratorios. De éstos, se clasificaron como estudios unicéntricos 350 (83,7 %), mientras que 45 (10,8 %) eran multicéntricos y 22 (5,5 %) no publicaron esta información. Los datos sobre los promotores del estudio aparecieron únicamente en 73 ECC (17,5 %), destacando la industria farmacéutica con 49 (67,1 %), y las agencias de financiación de la investigación y las fundaciones para la investigación con 17 (23,3 %).

Los principales temas investigados en los ECC identificados fueron: hipertensión arterial (40 ECC, 9,6 %), asma (22 ECC, 5,3 %), enfermedad pulmonar obstructiva crónica (20 ECC, 4,8 %), úlcera gástrica y duodenal (16 ECC, 3,8 %), deshabitación tabáquica (13 ECC, 3,1 %), diabetes mellitus (12 ECC, 2,9 %), seguidos por cardiopatía isquémica (11 ECC, 2,6 %) y obesidad (10 ECC, 2,4 %).

La fase del ensayo clínico sólo se explicitó en la publicación en 12 ECC (2,9 %), seis en fase I, cinco en fase III y uno en fase IV. El número de individuos que participaron en los ECC fue el siguiente: 240 ECC (57,4 %) realizados con menos de 50 participantes; 130 ECC (31,1 %) tuvieron entre 51 y 200 participantes, y 30 ECC (7,2 %) más de 200 participantes.

Según los autores, la asignación a una de las intervenciones del ECC se realizó mediante aleatorización en 345 ECC (82,7 %), mientras que en los 72 restantes (17,2 %) no se citó si la asignación al tratamiento fue aleatoria, pero en estos casos se mencionaba el método doble ciego para enmascarar el tratamiento asignado. Sólo 37 estudios (8,9 %) fueron clasificados como ECDA (estudios que cumplían los cuatro criterios de elegibilidad y el/los autor/es usaban un método explícito y válido de asignación aleatoria), mientras que los restantes 380 (91,1 %) fueron clasificados como EEC (estudios con métodos de asignación no descritos explícitamente o con procedimientos "casi al azar").

El enmascaramiento se utilizó en 176 ECC (42,2 %): 154 (36,9 %) fueron doble ciego y 22 (5,2 %) simple ciego. Un total de 19 ECC (4,5 %) eran descritos explícitamente como abier-

tos, mientras que en los 222 ECC restantes (53,2 %) no se mencionó ningún mecanismo para enmascarar el estudio. En 383 ECC (91,8 %) se utilizó un diseño de grupos paralelos, es decir, los participantes fueron asignados a una de las intervenciones durante la totalidad del ensayo. En cambio, el diseño cruzado (*crossover*), intercambiando intervenciones, se utilizó en 34 ECC (8,2 %).

La intervención fue de tipo farmacológico (p. ej., cromoglicato disódico frente a placebo) en 329 ECC (78,9 %) y no farmacológica (p. ej., técnicas de educación grupal, diferentes tipos de jeringuillas, dietas o ejercicio físico) en 68 (16,3 %). Se comparó una intervención farmacológica con una no farmacológica (p. ej., consejo antitabáquico frente a chicle de nicotina o técnicas de relajación frente a amitriptilina) en 20 ECC (4,8 %). El placebo fue utilizado en 96 (29,2 %) de los ensayos en los que se utilizó intervención farmacológica. Se comparó una sola intervención frente a grupo control en 296 ECC (71 %), mientras que en el resto, 121 (29 %), participaron tres o más grupos.

El período de seguimiento de 244 (58,5 %) de los ECC analizados fue inferior a los 6 meses desde que se realizó la asignación aleatoria hasta que se acabó el período de seguimiento y en 76 (18,2 %) superó los 6 meses. No se informó sobre el tiempo de seguimiento en los restantes 97 ECC (23,3 %).

## Discusión

Las revisiones sistemáticas se elaboran a partir de los estudios individuales y, por esta razón, su identificación es un paso ineludible. Para evitar duplicaciones y redundancias innecesarias, es preciso que esta actividad se realice de manera coordinada. Para identificar todos los ECC y facilitar el trabajo a los que van a realizar revisiones sistemáticas o simplemente quieren consultar toda la evidencia experimental existente sobre un determinado tratamiento, en los últimos años se coordina en el seno de la Colaboración Cochrane un registro mundial de ensayos clínicos (*The Cochrane Controlled Trial Register*), cuyo fin es evitar duplicaciones y esfuerzos innecesarios<sup>11,13,14</sup>.

El Centro Cochrane español está intentando contribuir a este registro aportando los ensayos clínicos realizados en nuestro país. Con este objetivo, se ha priorizado la identificación de los ECC publicados en las revistas españolas, y entre éstas, las de medicina general e interna. Para desarrollar este trabajo, conviene destacar, en primer lugar, la enorme dificultad para disponer del denominador de publicaciones españolas. Sólo combinando la información del IME, del CINDOC

(C-17), los catálogos de publicaciones de diversas Comunidades Autónomas y las comunicaciones orales de diversas personas se ha podido disponer de una lista de publicaciones biomédicas españolas lo suficientemente exhaustiva. Que el número de éstas, 732 en total y 68 de medicina general e interna, sea tan elevado, obliga a plantear, a nuestro entender, la conveniencia de una reducción en el número de publicaciones mediante la fusión de algunas con objetivos parecidos o la eliminación de las de menor calidad. La dispersión de publicaciones implica la dispersión de los trabajos y la de los lectores potenciales y contribuye a favorecer siempre la publicación de los manuscritos de menor calidad. Probablemente sería preferible disponer de menos revistas pero mejor editadas, lo que redundaría en menos trabajos pero de mayor calidad<sup>15</sup>. Se trata de que las publicaciones biomédicas no sean en ningún caso un mero receptáculo pasivo de trabajos y ejerzan el papel de crítica científica y validación social que les corresponde<sup>16,17</sup>.

Otra consecuencia del elevado número de revistas, pero no exclusivamente determinada por este factor, es la dificultad para poder acceder a las bibliotecas y a las revistas de manera suficientemente ágil. La existencia del IME, CINDOC y de los catálogos autonómicos supone sin duda una ayuda valiosa, pero nuestro país está aún lejos de disponer de redes de información y servicios documentales suficientemente completas y eficientes. Sólo gracias al espíritu de cooperación que ha animado a nuestros colaboradores, o a los documentalistas de las bibliotecas consultadas, entre otros, se pueden salvar estos obstáculos.

Nuestro trabajo ha constatado, una vez más<sup>18</sup>, las limitaciones de la búsqueda documental realizada exclusivamente a través de Medline, ya que esto comporta la pérdida de, al menos, una tercera parte de los ensayos clínicos españoles en lo concerniente a las revistas de medicina general e interna, con una proporción de falsos negativos del 32 % y de falsos positivos del 96 %. Nuestra proporción de pérdidas de ensayos es ligeramente menor que la encontrada en otros estudios<sup>19,20</sup> y se podría explicar, al menos en parte, por el escaso número de revistas españolas que están indizadas en Medline y que, por lo tanto, han podido ser incluidas en la comparación entre la búsqueda manual y la electrónica. Únicamente 6 de las 68 revistas españolas de medicina general e interna están o han estado indizadas en Medline, entre las 45 españolas de cualquier especialidad. Aunque la inclusión de una revista en Medline no es condición suficiente de excelencia, sin duda se trata de un referente necesario que debe tenerse

en cuenta. Que sólo el 6 % del total de revistas españolas y también del ámbito de la medicina general e interna estén indizadas en Medline refuerza la hipótesis apuntada de que en nuestro país existen excesivas publicaciones biomédicas.

En el período de 25 años analizado (1971-1995) se publicaron 417 ECC, es decir, a razón de unos 16 por año, aunque la tendencia de los últimos años ha sido mayor (fig. 1). Sólo una tercera parte (37 %), aproximadamente, de las revistas españolas han publicado al menos un ECC, siete de ellas sin estar indizadas en alguna base de datos y sólo seis lo han estado en Medline. MEDICINA CLÍNICA ha sido la publicación que más ECC ha incluido, con 93 (22 %) en todo el período, lo que supone menos de 4 ECC al año. Teniendo en cuenta que prácticamente todas las revistas dedican una buena parte de su atención a temas terapéuticos, el bajo número de ensayos clínicos publicados refleja, a nuestro entender, el insuficiente desarrollo de esta importante línea de investigación clínica en España, aspecto muy preocupante por todo lo que supone.

En cuanto a las características de los ECC identificados, sin duda el denominador común de la gran mayoría es la ausencia de datos y detalles básicos como son el período de realización del estudio, las fuentes de patrocinio o de financiación, la fase clínica del ECC, la explicitación de los mecanismos de aleatorización y enmascaramiento utilizados. El análisis de otras variables mejor informadas permite conocer que la hipertensión ha sido, con 40 ECC (10 %), el tema más investigado en España mediante este tipo de diseño; que, mayoritariamente, más de la mitad de estos estudios incluyeron menos de 50 individuos y una tercera parte, aproximadamente, entre 51 y 200 participantes; que en la gran mayoría, el diseño predominante fue el de grupos paralelos, utilizados para comparar un tratamiento farmacológico con un grupo control sin placebo; que el período de seguimiento ha sido inferior a los 6 meses; que la mayor parte de los ECC se realiza en Cataluña y en Madrid, y que, en conjunto, en muy pocos interviene la atención primaria.

Aunque la perspectiva de lo que es la investigación experimental clínica que se ha desarrollado en España no se completará hasta disponer de la información sobre los ECC españoles publicados en revistas de ámbito internacional y, sobre todo, de los realizados pero no publicados (estudio que permitirá estimar la magnitud del sesgo de publicación), ya se puede afirmar que, puesto que muy previsiblemente los ECC publicados tienen una mayor calidad que los no publicados y dado el gran número de insuficiencias

detectadas en aquéllos, el panorama general no es demasiado halagüeño<sup>21</sup>. Si bien nunca hay que confundir la mala información sobre los ECC que contienen los artículos que los refieren con la mala calidad intrínseca de los estudios, ambas están indisolublemente ligadas, ya que un estudio, por excelente que sea, pero que no se publica o se difunde de manera inadecuada, se convierte en un mal estudio al no cerrar el ciclo de la investigación científica<sup>22</sup> y, además, impide o dificulta enormemente la utilización de sus datos en cualquier revisión sistemática. De ahí la importancia de estrategias preventivas como es en la práctica la iniciativa CONSORT<sup>23</sup>; se trata de conseguir que las revistas se conviertan en un filtro de primer orden para evitar la publicación de ensayos si no se facilitan todos los detalles necesarios para su oportuna valoración. Sería muy deseable que las revistas españolas adoptaran de manera explícita, como han hecho otras publicaciones muy destacadas de otros países, el compromiso de publicar sólo aquellos ECC que se atengan a los requerimientos de la iniciativa CONSORT<sup>8</sup>. Además de los lectores, los propios autores se beneficiarían de recibir una retroalimentación (*feedback*) que les obligara a generar y proporcionar los detalles necesarios para la publicación de su ensayo clínico. Todavía mejor: mejorarían el diseño del estudio antes de su inicio, sin duda el mejor momento para prevenir posteriores limitaciones, problemas y críticas.

## BIBLIOGRAFÍA

- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *Br Med J* 1994; 309: 1.286-1.291.
- Lefebvre C. The Cochrane Collaboration: the role of the UK Cochrane Centre in identifying the evidence. *Health Libraries Review* 1994; 2: 235-242.
- Mulrow CD, Oxman AD, editores. *Cochrane Collaboration Handbook* [updated September 1997]. En: *The Cochrane Library* [database on disk and CDROM]. The Cochrane Collaboration. Oxford: Update Software, 1994, issue 4.
- Bonfill X. La medicina basada en la evidencia. En: Del Llano J, Ortún V, Martín JM, Millán J, Gené J, editores. *Gestión Sanitaria. Innovaciones y desafíos*. Barcelona: Masson, 1998; 487-512.
- Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Statist Soc A* 1988; 151: 445-463.
- Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Control Clin Trials* 1987; 8: 343-353.
- Moher D, Fortin P, Jadad AR et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; 347: 363-366.
- Urrutia G, Bonfill X, Martí J. En busca de la evidencia perdida. Algunas propuestas para incrementar la transparencia de los ensayos clínicos. *Med Clin (Barc)* 1999; 112 (Supl 1): 21-27
- Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* 1993; 2: 39 paragraphs.
- Jadad AR, Jamal S, Carroll D, McQuay HJ. Can volunteers with no research experience hand search medical journal effectively to find randomized controlled trials? *Control Clin Trials* 1994; 15: 62-73.
- The Cochrane Database of Systematic Reviews. [Update 1998]. Available in the Cochrane Library [Database on disk and CDROM]. The Cochrane Collaboration: Issue 2. Oxford: Update Software: 1998 [Update quarterly].
- Bonfill X, Lacalle JR, Avendaño C, Alexandre R, García-López F, García-Alonso F. Identificación y descripción de los ensayos clínicos realizados en España (1971-1995). *Fis N°: 96/1802E*.
- Bonfill X, Etcheverry C, Martí J, Glutting JP, Urrutia G, Pladevall M. El desarrollo de la Colaboración Cochrane española. *Med Clin (Barc)* 1999; 112 (Supl 1): 17-20.
- Dickersin K, Larson K. Establishing and maintaining an international register of RCTs. En: *The Cochrane Collaboration Handbook*. Oxford: The Cochrane Collaboration, 1995 [updated 14 July 1995]. Available in: *The Cochrane Library* [database on disk and CDROM]. Oxford: Update Software, 1996 [updated quarterly].
- International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *N Engl J Med* 1997; 336: 309-315.
- Munro AJ. Publishing the findings of clinical research. *Br Med J* 1993; 307: 1.340-1.341.
- Pearm J. Publication: an ethical imperative. *Br Med J* 1995; 310: 1.313-1.315.
- McDonald SJ, Lefebvre C, Clarke MJ. Identifying reports of controlled trials in the BMJ and the *Lancet*. *Br Med J* 1993; 313: 1.116-1.117.
- Jadad AR, Carroll D, Moore A, McQuay H. Developing a database of published reports of randomised clinical trials in pain research. *Pain* 1996; 66: 239-246.
- Adams CE, Power A, Frederick K, Lefebvre C. An investigation of the adequacy of MEDLINE searches for randomized controlled trials (RCTs) of the effects of mental health care. *Psychol Med* 1994; 24: 741-748.
- Soto J, Galende I, Sacristán JA. Calidad de los ensayos clínicos publicados en España: valoración a través del análisis de 3 revistas durante el período 1985-1991. *Med Clin (Barc)* 1994; 102: 241-245.
- Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990; 263: 1.405-1.408.
- 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: 637-639.

---

# Índice cronológico

---

## 1990

---

J. MORENO Y B. GONZÁLEZ GABALDÓN	
Investigación en humanos	129

---

## 1991

---

R. DAL-RÉ	
Investigación clínica con fármacos: estudio de la influencia de la información sobre reacciones adversas en la obtención del consentimiento informado	155

---

## 1992

---

B. SPILKER	
Cómo elevar la calidad de los ensayos clínicos y de su publicación	207
S. LOCK	
La revisión de manuscritos	210
T.C. CHALMERS	
La disminución del sesgo en la publicación de los ensayos clínicos	214
S. ERILL	
La ética de la publicación: el caso de los resultados negativos	219
P. BROWN, MD	
Editoriales, revisiones y cartas: por quién, para quién y cuándo	222

---

## 1993

---

D. GRACIA	
Entre el ensayo clínico y la ética clínica: las buenas prácticas clínicas	134
P. SIMÓN LORDA Y L. CONCEIRO CARRO	
El consentimiento informado: teoría y práctica (I)	161
J.A. SACRISTÁN, J. SOTO E I. GALENDE DOMÍNGUEZ	
Evaluación crítica de ensayos clínicos	81
I. GALENDE DOMÍNGUEZ	
Problemas éticos de la utilización de seres humanos en investigación biomédica. Consecuencias sobre la prescripción de medicamentos	138

---

## 1994

---

I. GALENDE DOMÍNGUEZ, J.A. SACRISTÁN Y J. SOTO	
Cómo mejorar la calidad de los ensayos clínicos	92

---

## 1995

---

J. CAMÍ	
Conflicto de intereses e investigación clínica	144
M.M. LÓPEZ MARTÍN Y J.E. BATISTA-MIRANDA	
Comités éticos de investigación: una complicación necesaria	189

---

En orden cronológico ascendente según su fecha de publicación en *Medicina Clínica*.

J. GARCÍA PUIG, F. MATEOS ANTÓN, A. GIL AGUADO Y C. BARCINA SÁNCHEZ	
¿Puede conocerse el “doble ciego” antes de desvelar los códigos de identificación?	3

**1996**

M.C. CARRÉ LLOPIS, J. JIMÉNEZ VILLA, M. MARTÍN MATEO Y F. JANÉ CARRENCÀ	
La estadística en la investigación clínica de medicamentos. Estudio de artículos originales procedentes de centros españoles	47
C. RUBIO TERRÉS	
Diseño estadístico de ensayos clínicos	56
P. SIMÓN LORDA, I.M. BARRIO CANTALEJO Y L. CONCEIRO CARRO	
Legibilidad de los formularios escritos de consentimiento informado	168
J.M. LÓPEZ ARRIETA Y N. QIZILBASH	
La medicina basada en pruebas: revisiones sistemáticas. La Colaboración Cochrane	227

**1997**

M.C. CARRÉ LLOPIS Y J. JIMÉNEZ VILLA	
La investigación clínica de medicamentos en el ámbito extrahospitalario	6
C. VALLVÉ	
La buena práctica clínica y la caja de Pandora (y III). La fiabilidad de los datos	101
E. GUALLAR, F.J. JIMÉNEZ, F. GARCÍA-ALONSO Y O.M. BAKKE	
La regresión a la media en la investigación y práctica clínica	66
J.B. CABELLO LÓPEZ, V. ABRAIRA Y J. GÓMEZ GARCÍA	
El ensayo clínico para un solo paciente. Justificación, metodología y aportaciones bioéticas	8
F. GARCÍA-ALONSO, E. GUALLAR, O.M. BAKKE Y X. CARNÉ	
El placebo en ensayos clínicos con medicamentos	19

**1998**

J. SOTO	
Utilidad de la incorporación de análisis de evaluación económica en los protocolos de los ensayos clínicos	26
R. DAL-RÉ, R. ORTEGA Y J. ESPADA	
Eficiencia de los investigadores en la selección de los enfermos para ensayos clínicos: a propósito de un estudio multinacional	28
X. BADIA Y X. CARNÉ	
La evaluación de la calidad de vida en el contexto del ensayo clínico	33

**1999**

A. CASADO, L. PRIETO Y J. ALONSO	
El tamaño del efecto de la diferencia entre dos medias: ¿estadísticamente significativo o clínicamente relevante?	72
M.I. LUCENA, G. TOGNONI Y F. SÁNCHEZ DE LA CUESTA	
Comités éticos de investigación clínica: nuevas funciones para un nuevo escenario	195
J.P. ORDOVÁS BAINES, E. LÓPEZ BRIZ, E. URBIETA SANZ, R. TORREGROSA SÁNCHEZ Y N.V. JIMÉNEZ TORRES	
Análisis de las hojas de información al paciente para la obtención de su consentimiento informado en ensayos clínicos	179
G. URRUTIA, X. BONFILL Y J. MARTÍ	
<i>En busca de la evidencia perdida.</i> Propuestas para incrementar la transparencia de los ensayos clínicos	105
J. MARTÍ, X. BONFILL, G. URRUTIA, J.R. LACALLE Y R. BRAVO	
Identificación y descripción de los ensayos clínicos publicados en revistas españolas de medicina general e interna durante el período 1971-1995	235
F.J. GARCÍA LÓPEZ, S. GUTIÉRREZ BEZÓN, I. GALENDE DOMÍNGUEZ Y C. AVENDAÑO SOLÁ	
Evaluación de la calidad de los ensayos clínicos: justificación, utilidad e inconvenientes	115

---

# Monografías Dr. Antonio Esteve publicadas

---

1. El hospital de día y su repercusión en terapéutica (1985).
2. Problemas que se plantean en el tratamiento de infecciones graves por *S. aureus* (1986).
3. Contribución del biólogo a la farmacología en España (1987).
4. Un glosario para farmacólogos (1987).
5. Aspectos biológicos de los síndromes depresivos (1988).
6. Bases del tratamiento de las intoxicaciones agudas (1988).
7. Investigación básica y medicina clínica (1988).
8. Tratamiento de datos en farmacología (1989).
9. Perspectivas terapéuticas en la esclerosis múltiple (1989).
10. Biotecnología de aplicación farmacéutica (1991).
11. Metodología del ensayo clínico (1991).
12. Periodismo científico. Un simposio internacional (1991).
13. El ensayo clínico como tarea cooperativa (1992).
14. Terapéutica y calidad de vida (1993).
15. Investigación sobre el cáncer en España: de la biología molecular a la clínica (1994).
16. El tratamiento del dolor: del laboratorio a la clínica (1994).
17. Farmacología de los canales iónicos (1995).
18. Bases de datos en farmacología y terapéutica (1996).
19. Fármacos y conducción de vehículos (1996).
20. Traducción y lenguaje en medicina (1997).
21. Medicina y medios de comunicación. Traducción al español de una serie publicada en la revista *The Lancet* (1997).
22. Problemas y controversias en torno al ensayo clínico (1998).
23. Glosario de investigación clínica y epidemiológica (1998).
24. Transducción de señales como diana farmacológica (1999).
25. Investigación médico-farmacéutica en atención primaria. Una visión a través de las publicaciones de la REAP (1999).
26. Modelos experimentales en patología infecciosa (2000).
27. Diccionario de farmacología y temas afines (2000).
28. Educación sanitaria: información al paciente sobre los medicamentos (2000).